



தமிழ் இணையம் 2010 தமிழ் இணையம் 2010 தமிழ் இணையம் 2010 தமிழ் இணையம் 2010 தமிழ் இணையம் 2010 தமிழ் இணையம் 2010 தமிழ் இணையம் 2010 தமிழ் இணையம் 2010

தமிழ் இணையம் 2010 கோவை

TAMIL INTERNET 2010 - COIMBATORE.



மாநாட்டுக் கட்டுரைகள்
CONFERENCE PAPERS

தமிழ் இணைய மாநாடு 2010

கோவை, ஜூன் 23 - 27

கட்டுரைகள்

Conference papers

Tamil Internet - 2010

Coimbatore, June 23 - 27



CONTENTS

(1) கணினி வழி தமிழ் மொழி கற்றல் மற்றும் கற்பித்தல்	001
1. சிங்கப்பூரில் தமிழ்மொழி கற்றல்-கற்பித்தலில் கணினியின் பயன்பாடு ஆ ரா சிவகுமாரன், தேசிய கல்வி நிறுவனம், சிங்கப்பூர். A Ra Sivakumaran, National Institute of Education, Singapore	003
2. மடிக்கணினியில் கன்னித்தமிழ் - ஒரு கற்றல் அனுபவம். சம்பந்தம் மோகன், கிரசன்ட் மகளிர் பள்ளி, சிங்கப்பூர் Sambandam Mohan, Crescent Girls School, Singapore	008
3. கல்விக்கூடங்களில் தமிழ் உபயோகத்தினைத் தரப்படுத்துதல். வெ. இராமன், அனன்யா வித்யாலயா V.Raman, Ananya Vidyalaya	013
4. ஆசிரியர் கல்வியில் கற்றல் மற்றும் வளக் கட்டமைப்பு Teaching and Resource Building in Teacher Education சீதாலட்சுமி, தேசிய கல்வி நிறுவனம், சிங்கப்பூர். Seetha Lakshmi, National Institute of Education, Singapore	019
5. அமெரிக்காவில் நடத்தப்படும் தமிழ்ப் பள்ளிக்கூடங்களில் நடைமுறையில் இருக்கும் தொழில்நுட்பத்தைப் பயன்படுத்துதல் Use of Technology in running a Tamil School in USA இளங்கோ மெய்யப்பன், கலிபோர்னியா தமிழ் நிறுவனம் Ilango Meyyappan, California Tamil Academy.	027
6. மின்னணு வழியில் தமிழ் மொழி கற்பித்தல் மற்றும் கற்றல். The E-learning and the teaching and learning of Tamil Language எம்.ஜே.ரபி சிங், மாநிலக் கல்லூரி, சென்னை, M. J. Rabi Singh, Presidency College.	033
7. தொடக்கக் கல்வி நிலையில் தமிழ் கற்பித்தல் மற்றும் கற்றல். Tamil Teaching and Learning at Elementary Level எல். நிர்மலா தேவி, கல்வி அரசு நிலையம் Nirmala Devi, Govt.Institute of Adv Study in Education	037
8. கணினி வழியில் தமிழ் கற்பித்தல் Teaching Tamil Language through Computer பரமசிவம் முத்துசாமி, புத்ரா பல்கலைக் கழகம், மலேசியா Paramasivam Muthusamy, Universiti Putra Malaysia	042
9. தமிழ் நாட்டில் தமிழ்க் கல்விக்கான தகவல் தொடர்பு நுட்பம் - நடைமுறைச் சவால்கள் ICT for Tamil Education in Tamil Nadu - Current Challenges எஸ் பாலாஜி, டி.பி.ஜெயின் கல்லூரி S.Balaji, D.B. Jain College	047
10. செயல்முறை அடிப்படையில் தமிழ்க் கற்றல் கற்பித்தலை நடைமுறைப்படுத்தல் மற்றும் இணைய ஒளிக்காட்சித் தரவுகள் வழி தமிழ் மொழிக் கற்றல் - தரவுகளை ஆய்ந்தறிந்து கற்கும் முறை. Enhancing Activity based Tamil Teaching and Learning Using Online Video Repositories: A Data Mining Based Approach. வி சரவணன், டாக்டர் என்.ஜி.பி தொழில் நுட்ப நிறுவனம், சென்னை. V. Saravanan, Dr. NGP Institute of Technology, Chennai.	052

11. கணினி - தமிழ் மொழி - தமிழ் ஆசிரியர் 056
The Computer, Tamil Language, The Tamil Teacher
ஜெய சரஸ்வதி துரைக்கண்ணு, பேடாக் மேற்கு தொடக்கப்பள்ளி, சிங்கப்பூர்.
Jayasarasvathi DuraiKannu, Bedok West Primary School, Singapore
12. தமிழ் கற்பித்தலில் கணினி மற்றும் இணையத்தின் பயன்பாடு 059
Using Computer and Internet resources in Tamil Teaching
கே. உத்தமன், ஆண்டர்சன் உயர்நிலைப்பள்ளி, சிங்கப்பூர்
K. Uthaman, Anderson Secondary School, Singapore.
13. வகுப்பறைக் கணினியில் தமிழ்க் கல்வி 063
Tamil Education through Classroom Computer
ந.வைரமணி, ராஜ ராஜேஸ்வரி பொறியியல் கல்லூரி
N.Vairamani, Rajarajeswari Engineering College
14. மழலைக்கல்வி - ஒர் ஆய்வு 067
Teaching Toddlers - An Analysis
பி.ஆர்.லக்ஷ்மி, அன்னைதேரசா மகளிர் பல்கலைக்கழகம், கொடைக்கானல்.
P.R.Lakshmi, Annai Therasa Women's University, Kodaikanal
15. தமிழ் மொழி கற்பித்தலில் கணினித் தொழில் நுட்ப பயன்பாடு 070
Uses of Computer Technology to Teach Tamil Language
கோவிந்தராஜ் ஏ, மீனாட்சிப்பல்கலைக்கழகம், சென்னை
Govindaraj. A, Meenakshi University, Chennai.
16. திறந்த நிலைக் கல்வி, தமிழ் முதல் மொழி 075
Open Education System, Tamil as the First Language.
எஸ் பால சுப்ரமணியன், பாரதியார் பல்கலைக்கழகம், கோவை.
N. Balasubramanian, Bharathiar University, Coimbatore
17. இங்கிலாந்தில் தமிழ்க் கல்வி கற்பிப்பதில் தகவல் தொடர்புத் தொழில் நுட்பத்தின் 080
பாவனை
Application of ICT in Teaching Tamil in England
சிவா பிள்ளை, லண்டன் Siva Pillai, London
18. அறிவு சார் கற்றல் நிலை 083
Knowledge based Learning framework
சங்கர் திலகவதி, போவன் பதின் உயர்நிலைப்பள்ளி, சிங்கப்பூர்
Sankar Thilavathi, Bowen Secondary School, Singapore
19. பல்லாடகப் பயன்பாட்டில் தமிழ்ப் பாடத்திட்டம் 088
எ மணவழகன், எஸ் ஆர் எம் பல்கலைக்கழகம் சென்னை
A. Manavazhahan, SRM University, Chennai
20. கணினி வழியில் தமிழ்ச் சொற்றொடர்களைக்கற்றல் 093
Computer Aided Learning in Tamil Sentences
ஜி.சிங்காரவேலு, பாரதியார் பல்கலைக்கழகம், கோவை
G.Singaravelu, Bharathiar University, Coimbatore

21.	கணினிவழி மாணவர்களிடத்தில் தமிழ்மொழியை வளர்த்தல் Developing Tamil Language Capability amongst Students through Computers சிவகௌரி கலியமூர்த்தி, பீக்கன் தொடக்கப்பள்ளி, சிங்கப்பூர். Sivagouri Kaliamurthy, Beacon Primary School, Singapore	098
(2)	இணையவழிக் கல்வி	105
1)	கூகுல் வழியாகத் தமிழ் கற்பித்தல் Teaching Tamil through Google முருகன், பெந்தமீர் உயர்நிலைப் பள்ளி, ஞானசேகரன், புனித தெரசா கான்வெண்ட், மோகன், கிரசண்ட் பெண்கள் உயர்நிலைப் பள்ளி Murugan, Bendemeer Sec Sch, Gnanasekaran, St. Theresa's Convent, Mohan, Crescent Girls Secondary School	107
2)	தமிழ் இணையப் பல்கலைக்கழக மென்பொருள்கள் - ஒரு கண்ணோட்டம் பி ஆர் நக்கீரன், தமிழ் இணையப் பல்கலைக்கழகம் P. R. Nakkeeran, Tamil Virtual University	113
3)	மூடுல் - மேம்பட்ட முறையில் கற்க, கற்பிக்க (தமிழ் மொழி) Moodle: For Enhanced Learning and Teaching of Tamil language ரவிஷங்கர் சோமசுந்தரம் Ravishankar Somasundaram	118
4)	தமிழ் இணையப் பல்கலைக்கழகம் - தரப்பகுப்பாய்வு Quality analysis of Tamil Virtual University எஸ்.ராஜ்குமார், குமரகுரு தொழில்நுட்பக் கல்லூரி S. Rajkumar, Kumaraguru College of Technology	123
5)	தொலைதூரத் தமிழ்க்கல்வியில் இணையம் வழிக் கற்றலின் பங்களிப்பு மன்னர் மன்னன் மருதை, மலேயா பல்கலைக்கழகம், மலேசியா Mannar Mannan Maruthai, University Malaya, Malaysia	127
6)	மூடுல் - தமிழ் கற்பித்தலுக்கான ஒரு பயன்பாட்டு மென்பொருள் Moodle: A Tool for Tamil Teaching கே. சர்வேஸ்வரன், இலங்கை K.Sarveswaran, Sri Lanka வி. நாகராஜன், இந்தியா V.Nagarajan, India.	134
(3)	கணினி மொழியியல்	139
1)	தமிழ் மொழிக்கான உருவியல் உருவாக்கி Morphological Generator for Tamil: A New Data Driven Approach எம். ஆனந்த் குமார், வி. தனலக்ஷ்மி, கே.பி. சோமன், அமிர்த விஸ்வ வித்யா பீடம், எஸ். ராஜேந்திரன், தமிழ்ப் பல்கலைக்கழகம் Anand kumar M, Dhanalakshmi V, Soman K.P., Amrita Vishwa Vidyapeedham Rajendran, S., Tamil University	141
2)	தேவாரத்தில் வினைச்சொல் திணை அமைப்புகள் Verbal Forms in the Tēvāram, in the light of the Tivākaram ழான்-லூய்க் செவ்வியார்-சி-என்-ஆர்-யெஸ் பல்கலைக்கழகம், பாரிஸ் Jean-Luc Chevillard, CNRS University, Paris	147

- 3) சங்க இலக்கியத் தொடரடைவி 155
கே உமாராஜ், மதுரை காமராசர் பல்கலைக்கழகம், மதுரை.
K. Umaraj, Madurai Kamaraj University, Madurai
- 4) தமிழ்த்தரவின் தன்மைகளும் கணிப்படுத்தலும் 159
Nature of Tamil Data and Computing
ப. டேவிட் பிரபாகர், சென்னை கிறித்துவக் கல்லூரி, சென்னை.
P.David Prabhakar, Madras Christian College
- 5) முக அசைவுகள் – தமிழ்ச் சொற்றொடரிலிருந்து முகத்தின் படம் கொணர்தல் 164
FaceWaves : A Tamil Text to Video Framework
மதன் கார்கி, டி.வி. கீதா, இரவி வர்மன், கிண்டி பொறியியல் கல்லூரி, அண்ணா பல்கலைக்கழகம்
Madhan Karky, T V Geetha, Ravi Varman, College of Engineering Guindy, Anna University
- 6) திருக்குறளில் சந்தர்ப்ப அடிப்படையில் தகவல் தேடல் 170
Context Based Information Search for Thirukural
ஜி.வி. உமா, என் இலக்கியச்செல்வன், கிண்டி பொறியியல் கல்லூரி, அண்ணா பல்கலைக்கழகம்
G.V. Uma, Ilakiyaselvan N, College of Engineering Guindy, Anna University
- 7) அசைநிலை உரை வடிவப் பாக்களுக்கான இடஞ்சாரா இலக்கணமும் பகுப்பாய்வு முறையும் 175
A Context-free Grammar and a Method to Parse Verses in Meter
ஈஸ்வர் ஸ்ரீதரன், இல. பாலசுந்தரராமன். Ishwar Sridharan, BalaSundaraRaman
- 8) தமிழைப் பல கோணத்தில் கற்க கணக்கீட்டு அணுகுமுறை 183
Computational Approaches for Learning Inflections in Tamil
கே ராஜன், முத்தையா பல் தொழில் நுட்ப கல்லூரி
K.Rajan, Muthiah Polytechnic College
- 9) இலக்கணப் பகுப்பான்கள் 190
Syntactic Parsers
மாரப்ப கணேசன், அண்ணா பல்கலைக்கழகம் Marappa Ganesan, Annamalai University.
- 10) கணினியில் தமிழும் பிற திராவிட மொழிகளும் 196
ராதா செல்லப்பன், பாரதிதாசன் பல்கலைக்கழகம்.
Radha Chellappan, Bharathidasan University
- 11) அறிவுசார் வெளிப்பாட்டிற்காக வரையறுக்கப்பட்ட அகராதி 203
Conceptual Lexicon for Knowledge Representation
எஸ் ராஜேந்திரன், தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்.
S.Rajendran, Tamil University, Thanjavur
- 12) கணினித்தமிழ் வளர்ச்சிக்கான அடிப்படை ஆராய்ச்சிப் பணிகள் 211
ந. தெய்வ சுந்தரம், சென்னைப் பல்கலைக்கழகம்
N. Deiva Sundaram, Madras University.

- 13) பெயர் சொற்றொடர் பகுப்பான் - ஒட்டு மொழிக்கான முடிவு நிலை தானியங்கி
Noun Phrase Chunker Using Finite State Automata for an Agglutinative Language 218
விஜய் சுந்தர், ஏ-யூ-கே-பி-சி ஆராய்ச்சி மையம், அண்ணா பல்கலைக்கழகம்
Vijay Sundar Ram R, AU-KBC Research Centre, Anna University.
- 14) அணிமேசனில் சங்கத்தமிழ்க் கவிதைகள் 225
Animated Sangathamizh Poems
அருள் நடராசன், தமிழ் வளர்ச்சித்துறை
Arul Natarajan, Tamil Development Department
- 15) சொல்வலையத்தில் உறவுமுறைச் சொற்கள் 228
Representation of Kinship in WordNet
எஸ் அருள்மொழி, திராவிடர் பல்கலைக்கழகம்
S Arulmozi, Dravidian University.
- 16) கணினியியலில் தமிழ்ப் பயன்பாடு 235
திருவள்ளுவன், சென்னைப் பல்கலைக்கழகம்.
Thiruvalluvan. I, Madras University
- 17) உருபனியல் பகுப்பாய்வில் குறிப்புத் தொடர்களின் பங்கு 239
Role of Regular Expression in Morphological Analysis
ஆர். சண்முகம், சென்னைப் பல்கலைக்கழகம்
R. Shanmugam, Madras University
- 18) கணினிக் கலைச்சொல்லாக்கத்தில் புதுச்சொல் உருவாக்கம் 242
எஸ் சார்லஸ், புனித சூசையப்பர் கல்லூரி, திருச்சி.
Charles. L, St. Joseph's College, Trichy.
- 19) எழுத்துருமாற்றிப் பணிகள்- தமிழ் வழி ரோமன், ரோமன் வழி தமிழ் 248
Transliteration Schemes for Tamil to Roman and Roman to Tamil
எஸ் ஸ்ரீனிவாசன்- இந்திராகாந்தி அணு ஆராய்ச்சி நிலையம்.
S. Srinivasan, Indira Gandhi Centre for Atomic Research
- 20) கன்னித்தமிழிருந்து கணினித்தமிழுக்கு... : ஒரு கண்ணோட்டம் 255
Tamil in Computers, A viewpoint
ஆ குமரன், மைக்ரோஸாப்ட் ஆராய்ச்சி மையம், பெங்களூரு
A Kumaran, Microsoft Research Centre, Bengalooru
- 21) தமிழ்க் கவிதைகளின் அமைப்பை தானியங்கி முறையில் அறிந்துகொள்ளல் 260
Automated Identification of Grammatical Patterns for Tamil Poems
எஸ் செந்தில் குமார், ஜி.எஸ். மகாலட்சுமி, என். பிரகாஷ், அண்ணா பல்கலைக்கழகம்
S. Sendhilkumar, G.S. Mahalakshmi, N. Prakash, Anna University
- (4) கணினி வழி தமிழ் மொழி சொந்திருத்திகள் 267**
- 1) தானியங்கு சொற்பிழைதிருத்தி உருவாக்கத்தில் உருபன் பகுப்பியின் பங்கு 269
ம. பார்கவி, கி. உமா தேவி, சென்னைப் பல்கலைக்கழகம்
- 2) தானியங்கு சந்திப்பிழைத் திருத்தி உருவாக்கத்தில் உருபன் பகுப்பியின் பங்கு 273
R. பத்ம மாலா, M. பார்கவி, சென்னைப் பல்கலைக்கழகம்

- 3) தானியங்கு சந்திப்பிழைத் திருத்தி உருவாக்கத்தில் முடிவு நிலை தானியங்கியின் பங்கு 277
Spell Checker for Tamil using Finite State Automata
அனிதா, எஸ். பிள்ளை, இந்துஸ்தான் பல்கலைக்கழகம் Anitha. S, Pillai, Hindustan University
- 4) தமிழுக்கான மொழியாராய்ச்சி மென்பொருள்கள்: கணினிமயமாக்கலுக்கான 282
தேவைகளும் தரப்படுத்தலின் முக்கியமும்
Language Tools for Tamil - Standardization for Computerization
எஸ் பாஸ்கரன், தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்.
S.Baskaran, Tamil University, Thanjavur.
- 5) மக்கள் தொகைக் கணக்கெடுப்பிணைப்படிவத்தில் தமிழில் தானியங்கு ஆய்வு 288
Automated Processing of Census Forms in Tamil
சுவப்பினில் பெல்ஹீ, சிடாக், ஷஷி கிரண், ரித்து ராஜ், சுரேஸ் சுந்தரம், ஏ.ஜி ராமகிருஷ்ணன்,
இந்திய அறிவியல் நிறுவனம்
Swapnil Belhe, CDAC, Shashi Kiran, Rituraj, Suresh Sundaram, AG Ramakrishnan
Indian Institute of Science
- 6) ஆங்கிலம்-தமிழ் மொழிபெயர்ப்புப் பொறி 295
Pattern Based English-Tamil Machine Translation
சரவணன், அமிர்த விஸ்வ வித்தியா பீடம். Saravanan, Amrita Vishwa Vidyapeetham.
- (5) கணினியில் தமிழ்ப் பேச்சு மற்றும் சொற்பகுப்பு ஆய்வு 301**
- 1) தமிழ் மற்றும் ஆங்கிலத்திற்கான இருமொழி தமிழ் உரையிலிருந்து பேச்சொலி மாற்றம் 303
Bilingual TTS for Tamil and English
ஏ.ஜி ராமகிருஷ்ணன், இந்திய அறிவியல் நிறுவனம்
A G Ramakrishnan, Indian Institute of Science
- 2) ஒலி அடிப்படையில் தமிழ் கற்பிக்கும் அமைப்பு 306
Voice Based Tamil Tutoring System
பி வாசுகி, எஸ் எஸ் என் பொறியியல் கல்லூரி, சென்னை.
P. Vasuki, SSN College of Engineering, Chennai.
- 3) தமிழுக்கான சொற்கள் இடைவெளி அறியும் உரையிலிருந்து பேச்சொலி மாற்றும் 314
மென்பொருள்
Prediction of Pauses in TTS – Tamil
பி. அருள் மொழி, இந்திய அறிவியல் கழகம் P.Arulmozhi, Indian Institute of Science
- 4) தமிழ் ஒலியன்கள் மற்றும் பேச்சுப் பகுப்பான் கொண்டு உரை பகுப்பாய்வு மற்றும் செய்தி 321
வெளிவாங்கு அமைப்பு
Text analysing & retrieval system using Tamil phonemes & VSM
ஆர் பிரேமலதா, ஆறுபடை வீடு தொழில் நுட்ப நிலையம்
Premalatha.R, Aarupadai Veedu Institute of Technology
- 5) வாய்பேச இயலாதோர்க்கான தமிழ்ப் பேசு பொறி 327
Speech Tool in Tamil for Speech Disabled
சஷி கிரண், இந்திய அறிவியல் நிறுவனம் Shashi Kiran, Indian Institute of Science

- 6) கணிணித் தமிழ் ஆய்வில் ஒலி பெயர்த்தல் முறைக்கான தரப்படுத்தல் 330
Standardisation of Modern Standard Tamil Transcription for Computational Tamil
புனல் முருகையன், அண்ணாமலைப் பல்கலைக்கழகம்.
Punal K. Murugaiyan, Annamalai University
- 7) உரை - பேச்சொலி மாற்றி - தமிழ் அசைகளை உணரும் மென்பொருள் 338
Realization of Tamil Syllables - Text To Speech Transferring System Using FPGA
டி. ஜெயசங்கர், அண்ணா பல்கலைக்கழகம், திருச்சி, அற்புத விஜய செல்வி, கிங்ஸ்
பொறியியல் கல்லூரி, புதுக்கோட்டை, ஆர். இராஜேந்திரன், தமிழ்ப் பல்கலைக்கழகம்
T. Jayasankar, Anna University, Tiruchirappalli, J. Arputha Vijaya Selvi, Kings
College of Eng. Pudukkottai, R. Rajendran, Tamil University, Thanjavur.
- 8) தமிழ் உரைகளை உதட்டசைவு மூலம் ஒலிக்கு மாற்றும் மென்பொருள் 345
FaceWaves : Tamil Text-To-Speech with Lip Synchronisation
தமிழரசன், அண்ணா பல்கலைக்கழகம்.
Tamilarasan, Anna University
- 9) தமிழ் உரைக்கான இரகசியம் காக்கும் முறைகள் 350
Classical Encryption Techniques for Tamil Text
பி நவனீதம், பி எஸ் ஜி தொழில் நுட்பக்கல்லூரி, கோவை
P.Navaneethan, PSG College of Technology, Coimbatore
- 10) தமிழ் உரைக்கோவைக்கான புள்ளியியல் பகுப்பாய்வும் தோற்றப்படுத்துதலும் 357
Statistical Analysis & Visualization of Tamil in Text Streams
ஜெ ஜெய் ஹரி ராஜு, கிண்டி பொறியியல் கல்லூரி, அண்ணா பல்கலைக்கழகம்
J. Jai Hari Raju, College of Engineering Guindy, Anna University
- 11) தமிழ் எழுத்துப் பிரதி எடுப்பதில் ஏற்படும் பல்வேறு முரண்பாடுகளைப் பற்றிய ஆய்வு 362
An Analysis of Various Types of Distortions of Tamil Scripts
K. அய்யாக்குட்டி, C ஜோதி வெங்கடேஸ்வரன், ஆர் இந்திரா காந்தி,
அன்னை தெரசா மகளிர் பல்கலைக்கழகம்.
K. Iyyakutti, C. Jothi Venkateswaran, R Indra Gandhi,
Mother Teresa Women's University
- 12) தமிழிற்கான அதி துல்லிய தொலைபேசிக் குரல் அங்கீகரிக்கும் அமைப்பு 372
A High Accuracy Phone Recognition System for Tamil
சி பி சந்தோஷ்குமார், சென்னைப் பல்கலைக்கழகம், என் தெய்வசுந்தரம்,
அமிர்த விஸ்வ வித்யா பீடம், கோவை.
C. P. Santhosh Kumar, Madras University, N. Deiva Sundaram,
Amrita Vishwa Vidyapeetham, Coimbatore.
- 13) தமிழ் உணர்வுச் சொற்கள் அடிப்படையில் முக வெளிப்பாடு 378
Face Waves: 2D Facial Expressions Based on Tamil Emotion Descriptors
சபிதா தம்மனேனி, மதன் கார்கி கணிணி, கிண்டி பொறியியல் கல்லூரி, அண்ணா
பல்கலைக்கழகம்.
Sabitha Tammaneni, Madhan Karky, College of Engineering Guindy, Anna University

(6) தமிழ் மின் தரவு மற்றும் மின்னகராதிகள்	385
1) மின்னணு நூலகம், மலேயாப் பல்கலைக்கழகம் E-Library of University of Malaya இளங்குமரன் Ilangkumaran	387
2) தமிழ் மின்னணு நூலகத்தின் முன்னேற்றம் Development of Tamil Digital Libraries கே.கல்யாணசுந்தரம், சுவிஸ் தொழில் நுட்பக் கழகம் K. Kalyanasundaram, Swiss Federal Institute of Technology	392
3) தரவுகளைப் பகுப்பாய்தல் மூலம் தமிழகத்தின் டிஜிட்டல் பிளவைக் குறைத்தல் Reducing Digital Divide in Tamilnadu using Data Mining Technology ஆர். ஜெய பிரபு, காருண்யா பல்கலைக்கழகம் R. Jayabrabu, Karunya University	399
4) தமிழ் மரபு சார்ந்த தகவல்களின் தகவல் வங்கி, மின் நூல்கள், ஓலைச்சுவடிகளின் ஒருங்கிணைக்கப்பட்ட இணைய அட்டவணை சுபாஷினி டிரம்மல், ஹூல்ட்-பாக்கார்ட், ஜெர்மனி Subashini Tremmel, Hewlett-Packard, Germany	406
5) தமிழ் நாடு விவசாயிகளுக்காக விவசாய தளங்களில் நீடிக்கப்பட்ட குறுக்கீட்டு மொழி தகவல் அளிப்பு முறையும் புள்ளி விவர ஆவன மொழிபெயர்ப்பும் An Extended Cross Lingual Information Retrieval System for Agricultural Domain Using Statistical Document Translation for Tamil Farmers டி.தேன்மொழி குழுவினர், எஸ்எஸ்என் பொறியியல் கல்லூரி, சென்னை D. Thenmozhi et al. SSN College of Engineering, Chennai	410
6) தமிழ் மின்னணு பெட்டக மேலாண்மை Tamil E-Archives Management பி. அசதுல்லா, ம. உ. கல்லூரி, ஆம்பூர். B. Asadullah, Mazharul Uloom College, Ambur.	415
7) தமிழ் ஆவணங்களுக்கான நீண்ட கால பராமரிப்பு வழி முறைகள் Long-term Preservation and Archival Strategies for Tamil Documents மணி மு. மணிவண்ணன், சிமாண்டெக் நிறுவனம். Mani M. Manivannan, Symantec Corporation.	419
8) இணையத்தில் தமிழ் மின் அகராதிகள்: ஒரு பார்வை A view of the Tamil Dictionaries in the Internet மணியரசன் முனியாண்டி, மலேசியா. Maniyarasan Muniandy, Malaysia	428
9) இணையத்தில் தமிழ் நூல்கள் Tamil Books in the Internet க.துரையரசன், அரசினர் கலைக் கல்லூரி, கும்பகோணம். K. Durairasan, Government Arts College, Kumbakonam	442
10) மின்னணு அருங்காட்சியகம் Electronic Museum மறைமலை இலக்குவனார், மாநிலக் கல்லூரி, சென்னை Maraimalai Ilakkuvanar, Presidency college, Chennai	447

11) தமிழ் விக்கிபீடியா எனும் தமிழ்க் கலைக்களஞ்சியம் Tamil Wikipedia தேனி சுப்ரமணி Theni.M.Subramani, www.muthukamalam.com	452
12) தமிழில் கணினிவழிச் சொல்லடைவு Indexing in Tamil through Computers இல. சுந்தரம், பாரதிதாசன் பல்கலைக்கழகம், திருச்சிராப்பள்ளி. L. Sundaram, Bharathidasan University, Tiruchirapalli	458
(7) இணையத்தொழில்நுட்பத்தில் தமிழ் மொழி மற்றும் திறவூற்று செயலிகள்	465
1) தமிழ் இணையச் சமூகங்களை ஊக்குவித்து, தமிழ் இணையப் பக்கங்களை ஆராய்தல் Fostering Tamil Web Communities for mining Tamil Web Pages விஜய கதிரவன். கே.எஸ்.ஆர் தொழில்நுட்பக் கல்லூரி, திருச்செங்கோடு A. Vijaya Kathiravan, K.S.R. College of Technology, Tiruchengodu.	467
2) தமிழில் பன்னாட்டு இணைய வலை முகவரி (ஐ-டி-என்) Tamil IDN எஸ் மணியம் S.Maniam, www.i-Dns.Net	475
3) மின்னணு தமிழ் அகராதியுடனான தமிழ் இணையத்தள உருவாக்கம் A Tamil Web Portal Development with online Dictionary எ .பரிமளா தேவி, குமரகுரு தொழில்நுட்பக்கல்லூரி. A Parimaladevi, Kumaraguru college of Technology.	480
4) தமிழ் மொழி காணொலிக் கருத்தரங்கு -2010 – வழிமுறைகள் Methods and Options of Videoconferencing in relation to the Tamil Language in 2010, எரிக் மில்லர், பென்சில்வேனியா பல்கலைக்கழகம் Eric Miller, University of Pennsylvania.	484
5) மலேசியத் தமிழ்ப்பள்ளிகளில் வகுப்பறைகளை நிர்வகிக்க ஒரு கணினி முறை Thin Client and Server Based Computing to Provide Integrated School and Class Room Management System in Malaysian Tamil Schools சாமிநாத குமரன் வேலு, சரவணன் மாரியப்பன், நெக்சஸ் தகவல் தொழில் நுட்ப தீர்வகம் Saminatha Kumaran Veloo, Saravanan Mariappan, Nexus IT Solutions, Malaysia.	490
(8) கணினி வழி தமிழ் எழுத்து உணரி செயற்பாடுகள்	495
1) எழுத்துணரிகளுக்கு வசதியான எழுத்துருக்களின் அமைப்பு முறை Embedding with optical co-features to 'OCR-Friendly' Fonts லோக சுந்தரம். ND Logasundaram.	497
2) தமிழ்ச் சொற்களை உணரும் மென்பொருளுக்கான குறிக்கப்பட்ட கையெழுத்து உரைத் தரவு உருவாக்கல் Creation of Annotated Tamil Handwritten Word Corpus for OCR நேத்ரவதி பி., இந்திய அறிவியல் கழகம் Nethravathi .B, Indian Institute of Science	502
3) தமிழ் எழுத்துணரிகளின் தயாரிப்பில் உள்ள சிக்கல்கள் Difficulties in Developing OCR for Tamil Documents ஆர். திவ்யபாரதி, பன்னாரி அம்மான் தொழில்நுட்பக் கழகம் R. DhivyaBharathi, Bannari Amman Institute of Technology	507

- 4) பழங்கால தமிழ் எழுத்துகளை உணரும் மென்பொருள் 514
 Recognition of Ancient Tamil Characters
 எஸ். இராஜக்குமார் டி.எம்.ஐ. பொறியியல் கல்லூரி
 S.Rajakumar, DMI College of Engineering
- 5) தமிழ் எழுத்துணரிகளை வடிவமைத்தலும் திறன் அறிதலும் 519
 Design and Evaluation of Omnifont Tamil OCR
 துஷ்கர் பட்டாநாயக், சிடாக், நோய்டா Tushar Patnaik, CDAC, Noida
- 6) தமிழ் கையெழுத்தை உணரும் மென்பொருள் வடிவமைப்பு வழிவகைகள் 523
 Approach to Recognize Handwritten Tamil Characters
 என். சாந்தி. கே.எஸ். ரங்கசாமி தொழில்நுட்பக் கல்லூரி
 N. Shanthy, K. S. Rangasamy College of Technology
- 7) கையடக்கக் கணினிகளில் தமிழ் எழுத்துணரிகளை அமைப்பு உணரல் முறையில் 531
 உருவாக்கல்
 Effective Tamil Character Recognition in Tablet PCs Using Pattern Recognition
 பெர்டின் ஜோ ஜே., ஆர் வேலாயுதம், ஐன்ஸ்டீன் பொறியியல் கல்லூரி,
 டி. இரவி, கே. சி. ஜி. தொழில் நுட்பக் கல்லூரி
 Ferdin Joe J, R. Velayutham, Einstein College of Engineering,
 T. Ravi, KCG College of Technology
- (9) தமிழில் சிந்தனைத்திறன் கணினிச் செய்நிரல்கள் 539**
- 1) அறிவுத் திறன் வாய்ந்த தமிழெழுதும் இயந்திரக் கை வடிவமைப்பு 541
 Design of Intelligent Robotic Arm Writing Tamil
 செய்மா சகி எஸ்., ஸ்ரீ கிருஷ்ணா தொழில் நுட்பக் கல்லூரி
 Seima Saki S, Sri Krishna College of Engg & Tech
- 2) லாலலா - தமிழ்ப் பாடல் பகுப்பாய்வு மற்றும் பாடல் எழுதும் மென்பொருள் 547
 LaaLaLaa - A Tamil Lyric Analysis & Generation Framework
 சௌமியா டி., அண்ணா பல்கலைக்கழகம் Sowmiya. D, Anna University
- 3) இயற்கை மொழிப் பகுப்பாய்வுக் மென்பொருள்கள் கொண்டு 553
 இணையக் கட்டுரைகளில் மனித உணர்வு வெளிப்பாடுகளை அறிதல்
 Language Independent Emotion Recognition System for
 Web Articles using NLP Techniques
 டி. மகேந்திரன், எஸ் குணசுந்தரி, பி. ராஜலட்சுமி, வேலம்மாள் பொறியியல் கல்லூரி
 D. Mahendran, S. Gunasundari, B. Rajalakshmi, Velammal Engineering College
- 4) அறிவுத்திறன் வாய்ந்த தன்னிச்சையாகத் தமிழ்க் கதை எழுதும் மென்பொருள் 559
 An Integrated Intelligent Framework for Automatic Story Generation in Tamil
 ஜி.வி. உமா. அண்ணா பல்கலைக்கழகம் G.V. Uma, Anna University
- 5) தமிழ் உரைகளைப் பகுத்தாயும் மென்பொருள் 567
 Evaluation of Tamil Descriptive Passages using Concept Maps
 மகாலட்சுமி ஜி.எஸ்., செந்தில்குமார் எஸ்., ஷங்கர் பி. அண்ணா பல்கலைக் கழகம்
 Mahalakshmi G.S., Sendhilkumar S., Shankar B., Anna University

- 6) தமிழ் உரையிலிருந்து கையசைவு மொழிக்கு மாற்றும் மென்பொருள் 572
A Machine Translation System for Converting Tamil Text-to-Sign Language
டி. நரசிம்மன், டி. மாலா. அண்ணா பல்கலைக்கழகம்
D.Narashiman, T.Mala, Anna University
- 7) இயந்திர மொழிபெயர்ப்பிற்கான அகராதி உருவாக்கம் 579
Lexicon for Machine Translation
ப. குமார், சென்னைப் பல்கலைக்கழகம் P. Kumar, Madras University
- 8) தமிழ் மீ- இலக்கணம் 585
Tamil Hyper Grammar
உமா மகேஸ்வரராவ் ஜி, கிரிஸ்டோபர் எம், பரமேஸ்வரி கே.,
ஹைதராபாத் பல்கலைக்கழகம்
Uma Maheshwar Rao G, Christopher M, Parameshwari K, University of Hyderabad
- 9) தமிழ் - தெலுங்கு இரு வழி மொழிபெயர்ப்பு மென்பொருள் 592
A Tamil - Telugu Bi-directional Machine Translation System
கிரிஸ்டோபர் எம், கிருபானந்தம் என், பரமேஸ்வரி கே, உமாமகேஸ்வர ராவ் ஜி,
விஜய பாரதி டி. ஹைதராபாத் பல்கலைக்கழகம்
Christopher M, Krupanandam N, Parameshwari K, Uma Maheshwar Rao G,
Vijaya Bharathi D, University of Hyderabad
- 10) தெலுங்கு - தமிழ் இருவழி மொழிபெயர்ப்பு மென்பொருளுக்கான அகராதி 601
தயாரித்தலில் உள்ள சில சிக்கல்கள்
Some issues in the Development of Telugu - Tamil Bilingual Lexicon for
Machine Translation
பரமேஸ்வரி கே, லாவன்யா ஜே, கிருபானந்தம் என், உமா மகேஸ்வர ராவ் ஜி.
ஹைதராபாத் பல்கலைக்கழகம்
Parameshwari K, Lavanya J, Krupanandam N, Uma Maheshwar Rao G,
Christopher M. University of Hyderabad.
- 11) இ.ஐ.எல்.எம்.டி. இந்திய மொழிகளுக்கான ஒருங்கிணைந்த இயந்திர 609
மொழி பெயர்ப்புக் கண்ணோட்டம்
EILMT: A Pan-Indian Perspective in Machine Translation
ஹேமந்த் தர்பாரி, அனூராதா லீலே, அபூர்பா தாஸ்குப்தா, பிரியங்கா ஜெயின்,
சி-டாக் பூனா, சரவணன், அம்ரிதா பல்கலைக்கழகம்
Hemant Darbari, Anuradha Lele, Aparupa Dasgupta, Priyanka Jain, C-DAC, Pune,
Saravanan, Amrita University.
- (10) கணினியில் தமிழ்த் தட்டச்சு 621**
- 1) கருத்துருக்கள் - விசைப்பலகை அமைப்புகள் 623
ஏ.ஆர். அமைதி ஆனந்தம் - விரைவுச் சாலைத் துறை - தமிழக அரசு.
A.R.Amaithi Anantham, Highways Department, Government of Tamil Nadu

- 2) வணிக, சொந்தப் பயனுக்கான இலவசத் தமிழ்த் தட்டச்சு விசைப்பலகை இடைமுகம் 634
Free Tamil Keyboard Interface for Business and Personal Use
அபிநவ சிவகுமார், அசஷ்ய் ராவ், ஏ.ஜி. ராமகிருஷ்ணன் - இந்திய அறிவியல் கழகம் -
பெங்களூரு.
Abhinava Shivakumar, Akshay Rao, A.G. Ramakrishnan, Indian Institute of Science,
Bangalore
- 3) தமிழ்த் தட்டச்சு மென்பொருள்களின் வடிவமைப்பு மற்றும் வளர்ச்சி 641
Development and Evolution of Tamil Keyboard Input Systems
இரவீந்திரன் கே. பால் - மலேசியா Ravindran K. Paul, Malaysia
- (11) தமிழ் வலைப் பூக்கள் 649**
- 1) தமிழ் வளர்ச்சியில் வலைப்பூக்கள் 651
The role of Blogs in the Growth of Tamil Language
துரை மணிக்கண்டன் - டாக்டர் கலைஞர் கலை மற்றும் அறிவியல் கல்லூரி
Dr. Durai Manikandan, Dr.Kalaingar Arts & Science College
- 2) அதிகார மையங்களும் வலைப்பதிவுகளும் 656
Blogs and Centres of Power
என். இளங்கோ, பட்ட மேற்படிப்பு மையம், புதுச்சேரி
N.Ilango, Centre for Postgraduate Education, Puducherry
- 3) தமிழ் வலைப்பதிவுகள் - கருவிகள், திரட்டிகள் ஆகியவற்றைத் தாண்டி 661
Tamil Blogs - Tools, Aggregators and Beyond
காசி ஆறுமுகம் - Kasi Arumugam
- 4) வளர்ந்துவரும் மலேசியத் தமிழ் இணைய ஊடகம் 667
Growing Tamil Internet in Malaysia
சுப.நற்குணன், மலேசியா Suba. Nargunan, Malaysia
- (12) மின்னரசும் தமிழ் தகவல் தொழில்நுட்பமும் 673**
- 1) மின்னாளுமை இணையச் செயலகம் 675
E-Governance - Internet Secretariat
ஆல்பர்ட் பெர்னாண்டோ - Albert Fernando
- 2) RFID நுட்பம் மூலமாக மின் வணிக அரசு வழங்கும் மின் அட்டைக்கான வடிவமைப்பு 679
Prototype for E-government issued E-card using RFID technology
விஜயலட்சுமி எஸ். ஆர்., டாக்டர் ஜி.ஆர்.டி. அறிவியல் கல்லூரி
Vijayalakshmi.S.R, Dr. GRD College of Science
- 3) தமிழ்நாட்டில் மின் ஆட்சி - ஒரு வழிகாட்டுதல் 684
Recommended Approach for e-Governance in Tamil Nadu using MPMLA.IN.
சையத் ஹுசைன் - Syed Hussain, mpmla.in
- 4) சிங்கையில் நான் அனுபவித்த தமிழ்த் தொழில்நுட்ப வளர்ச்சி 689
Technology Growth in Tamil Experienced in Singapore
மீனாட்சி சபாபதி, சிங்கப்பூர் Meenatchi Sabapathy, Singapore

- 5) இலங்கையில் மின்னரசாங்கம் நோக்கிய தமிழ் தகவல் தொழில்நுட்பம் - சவால்களும் சாதனைகளும் 694
Tamil IT and E-Governance in Sri Lanka: Challenges and Achievements
தங்கராஜா தவரூபன், ஸ்பீட் ஐடி நெட், இலங்கை
Thangarajah Thavarupan, Speed IT net, Sri Lanka
- 6) தமிழகத்தில் மின்னரசு முயற்சிகள் 702
E-Governance Initiatives in Tamil Nadu
இ. இனிய நேரு, நடுவன் செய்தி மையம், சென்னை
E.Iniya Nehru, National Informatics Centre, Chennai
- 7) தமிழ் இணையப் பல்கலைக்கழகத்தின் தமிழ் மின்னாட்சிச் செயல்பாடுகள் 709
E- Governance in Tamil for Tamil Virtual University
ஜி. அமிர்தராஜ், ஏ. ஜேம்ஸ், பி.ஆர். நக்கீரன், தமிழ் இணையப் பல்கலைக்கழகம்
G. Amirtharaj, A. James and P R Nakkeeran, Tamil Virtual University.
- (13) கணினி வழி கல்வி 717**
- 1) கற்பிக்கவும் கற்கவும் ஆராய்ச்சி செய்யவும் மின்-வளங்களே சிறந்தவை 719
E-Resources are the best Information Service to Teach, Learn and Research
through World Wide Web
வி. தங்கவேல், டி.மோஹன்ராஜ், ரமேஷ் V. Thangavel, D. Mohanraj, Ramesh
- 2) தொடக்கப் பள்ளிக்கு தேவையான தமிழ்க் கணினி தேவைகள் 731
Computing Requirements for an Elementary School
ஆண்டோ பீட்டர் Anto Peter, Softview.in
- 3) மின் வழி தமிழ்க் கற்றலின் திறனாய்வு வழிவகைகள் 733
Quality Assessment Technique of E-Learning in Tamil Language
ஏ. கோவலன், பெரியார் மணியம்மை பல்கலைக்கழகம்
A. Kovalan, Periyar Maniammai University
- 4) தமிழ்க் கல்வியில் பல்லுடகத் தொழில்நுட்பத்தின் தாக்கம். 741
Impact of Multimedia technologies for Tamil education
பி. ஜெகதீசன், டி.பி. ஜெயின் கல்லூரி B. Jagadhesan, D. B. Jain college
- 5) இன்றைய மாணவர்களுக்கான தமிழ் கணிமை திட்டங்கள் - எங்கள் கல்லூரி அனுபவம் 744
Tamil Computing Projects for Today's Students - Our Experience
முத்துக்குமார் ஆறுமுகம் - குமரகுரு தொழில்நுட்பக் கல்லூரி
Muthukumar Arumugam, Kumaraguru College of Technology, India
- 6) தமிழ் முதனிலைக் கல்வியில் பல்லுடகப் பயன்பாட்டு முறை 747
Teaching Primary Education in Tamil Using LMS and Visualization Techniques
ரிச்சார்ட்ஸ் ஹாட்லி ஆர்., ஈ. இனிய நேரு Richards Hadlee. R, E. Iniya Nehru
- 7) கலைஞர் தமிழ்பேசும் கணினிக்கண் உருவாக்கத் திட்டம் 753
முனைவர் ஆர். ஜெயச்சந்திரன் - பாரதிதாசன் பல்கலைக்கழகம்
Dr. R. Jayachandran, Bharathidasan University

- 8) தமிழ் மொழி வளர்ச்சியில் கணினி நுட்பத்தின் பங்கு 757
 Role of Cloud Computing in Tamil Language Development
 ஆர். ராஜ ராஜேஸ்வரி, ஏ. பெத்தலட்சுமி, எம்.வி.எம். அரசுக் கல்லூரி, திண்டுக்கல்
 R. Raja Rajeswari, A. Pethalakshmi, M.V.M Government Arts College, Dindigul
- (14) தமிழில் தேடு பொறிகள் 761**
- 1) தமிழ் இணையத்தளமும் தேடு பொறிகளும் 763
 கபிலன் - மதுரை காமராசர் பல்கலைக்கழகக் கல்லூரி
 Kabilan, Madurai Kamaraj University College
- 2) இடைமுகத் தமிழ் உள்ளீட்டு மென்பொருட்கள் - ஓர் ஒப்பீடு 767
 பன்னிருகை வடிவேலன் ஆர்., சென்னைப் பல்கலைக்கழகம்.
 Pannirukaivadivelan. R, Madras University
- 3) தமிழ் மற்றும் தமிழல்லாத பயன்பாட்டாளர்களுக்கான செய்திக் கொணர் மென்பொருள் 772
 Information Retrieval System for Tamil and non-Tamil users
 எஸ். ஸ்ரீவித்யா, டி. மாலா, அண்ணா பல்கலைக்கழகம்
 S.Srividhya, T. Mala, Anna University, Chennai
- 4) தமிழில் தேடுபொறி! 776
 செல்வமுரளி
- 5) தேடல் - தமிழ் இணையத் தகவல்களைத் தேடும் தேடுபொறிகளின் மன்னர் 781
 Searchko- The King of Search for Tamil Web Documents
 ஷோப லலிதா தேவி - ஏயு-கேபிசி ஆராய்ச்சி மையம் - அண்ணா பல்கலைக்கழகம்.
 Sobha Lalitha Devi, AU-KBC Research Centre, Anna University
- 6) கோரி தேடல் அமைப்பு : பொருள் விரிவாக்க முறை 786
 இளஞ்செழியன், கிண்டி பொறியியல் கல்லூரி, அண்ணா பல்கலைக்கழகம்
 Ilanchezhian, College of Engineering Guindy, Anna University
- 7) கோரி - உயர்நிலை தேடல் மென்பொருள் 794
 CoRe - A Framework for Concept Relation Based Advanced Search Engine
 டி.வி. கீதா, ரஞ்சனி பார்த்தசாரதி, மதன் கர்க்கி, அண்ணா பல்கலைக்கழகம்
 T V Geetha, Ranjani Parthasarathi, Madhan Karky, Anna University
- (15) கையடக்கக் கருவிகளில் தமிழ் 801**
- 1) தமிழில் செல்லிட ஆளுகை 803
 அண்ணாக்கண்ணன், பச்சையப்பன் கல்லூரி
 Annakannan, Pachappas College
- 2) திருக்குறள் கையடக்கத் தொலைபேசி - அனைவருக்குமான ஒரு பண்பாட்டுக் கருவி 814
 Thirukkural Mobile: A cultural Tool for All
 ஜி. புவன் பாபு G. Bhuvan Babu, Avon Mobility Solutions
- 3) செல்பேசிகளில் தரப்படுத்தப்பட்ட தமிழ் இடைமுகம் 817
 எம். சிவலிங்கம் - பி.எஸ்.என்.எல் M. Sivalingam, BSNL

- 4) தமிழ் செல்பேசி வழியில் கல்விக்கான மென்பொருள் வடிவமைப்பு 822
Software Architectures for Tamil Mobile Learning
சுவர்ணலதா Swarnalatha, Apex Micro System
- 5) செல்பேசிகளில் தமிழ்க் குறுஞ்செய்தி முன்னுரைத்தல் 828
Predictive Tamil Short Messages for Handheld Devices
அபிராமி எஸ்., அண்ணா பல்கலைக்கழகம் Abirami. S, Anna University
- (16) தமிழ் ஒருங்குறி 831**
- 1) இந்திய அச்சு நிறுவனங்களில் ஒருங்குறியின் நிலை 833
Status of Unicode in the Indian Tamil Publishing Industry
பி. செல்லப்பன் - பழனியப்பா பிரதர்ஸ் P. Chellappan, Palaniappa Bros
- 2) மென்பொருள் கட்டமைப்பிலும் இயற்கை மொழி பகுப்பாய்வு 837
மென்பொருள்களிலும் ஒருங்குறியின் பயன்பாட்டில் சிக்கல்கள்
Problems of using Unicode in Software Components and in NLP
வி. கிருட்டிணமூர்த்தி - சென்னை V. Krishnamoorthy, Chennai
- 3) தமிழ் ஒருங்குறியை அச்சு நிறுவனங்களிலும் மின்-ஆளுமையிலும் 843
பயன்படுத்துவதில் ஏற்படும் சவால்கள்
Challenges to Publishing and E-Governance with Tamil Unicode
ஏ. இளங்கோவன் - கேட்கிராப் மின்மையத் தனியார் நிறுவனம்
A. Elangovan, Cadgraf Digitals Private Limited
- 4) மின்வழி ஊடகத்தில் தமிழ் எழுத்துருக்கள் - ஒரு ஆய்வு 850
A study on Tamil script in Digital Media
என். அன்பரசன், ஆப்பிள்சாப்ட்
N. Anbarasan, Applesoft

வாழ்த்துச் செய்திகள்

மு. கருணாநிதி
முதலமைச்சர்



தலைமைச் செயலகம்
சென்னை - 600 009

நாள்...16-06-2010

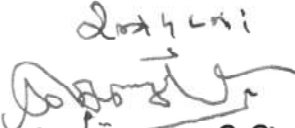
வாழ்த்துச் செய்தி

உலகத் தமிழ்ச் செம்மொழி மாநாட்டுடன் இணைந்து, உலகத் தமிழ் இணைய மாநாடும் கோவையில் நடைபெறுவதையொட்டிச் சிறப்பு மலர் வெளியிடப்படுவது அறிந்து மகிழ்கிறேன்.

தமிழ்த் தகவல் தொழில்நுட்பம் குறித்து இதுவரை, சென்னையில் 1999, 2003 ஆம் ஆண்டுகளிலும், சிங்கப்பூரில் 1997, 2000, 2004 ஆம் ஆண்டுகளிலும், மலேசியாவில் 2001 ஆம் ஆண்டிலும், அமெரிக்காவில் 2002 ஆம் ஆண்டிலும், செருமனியில் 2009 ஆம் ஆண்டிலுமாக எட்டு உலகத் தமிழ் இணைய மாநாடுகள் நடைபெற்றுள்ளன. அவற்றின் தொடர்ச்சியாகக் கோவையில் உலகத் தமிழ்ச் செம்மொழி மாநாட்டின் சிறகாக, உலகத் தமிழ் இணைய மாநாடு வடிவமைக்கப்பட்டுள்ளது. இம்மாநாடு, தமிழ்க் கணினியம், தமிழ் இணையம் ஆகியவற்றின் அண்மைக்கால முன்னேற்றங்கள், பெருமுயற்சிகள் குறித்த கருத்துகளைப் பரிமாறிக் கொள்ளவும், கால மலர்ச்சிக்கேற்ப எழும் புதிய சிந்தனைகளுக்கு வடிவம் தரும் தீர்வுகளை நோக்கி முன்னேறவும் வழிவகுக்கும் என நம்புகிறேன்.

காலத்தால் சாலப்பழைமையுடைய நம் செம்மொழியாம் அருமைத் தமிழ்மொழி பல்லாயிரம் ஆண்டுகளாகத் தொடர்ந்து வளர்ந்து சீரிளமைத் திறத்தோடு விளங்குவதுடன், கணினிச் சோலையில் மலரும் கண்கவர் மலராகவும் திகழ்கிறது.

அனைத்துத் தமிழ் இலக்கியங்களையும் அசைகலை வடிவில், இசையோடு இழைத்துத் தலைசிறந்த கலை இலக்கிய அறிவுப் பெட்டகமாக எதிர்காலத் தலைமுறையினருக்கு வழங்கிடும் திருப்பணிகளை நிறைவேற்றிடும் நோக்கில் வழிகாட்டி உலகத் தமிழ் இணைய மாநாடு வெற்றிபெற எனது உளமார்ந்த நல்வாழ்த்துகள் உரித்தாகுக.


(மு.கருணாநிதி)

மு.க. ஸ்டாலின்
துணை முதலமைச்சர்



தலைமைச் செயலகம்
சென்னை - 600 009

நாள் 16/6/2010

வாழ்த்து செய்தி

உயர் தனிச் செம்மொழியாம் நம் தமிழ் மொழியின் சிறப்பை போற்றிடும் வகையிலும், அதனை மேலும் நிலை பெறச் செய்திடத் தேவையான வழிமுறைகளைக் காணவும், மாண்புமிகு தமிழக முதலமைச்சர் கலைஞர் அவர்களின் எண்ணப்படி 'உலகத் தமிழ்ச் செம்மொழி மாநாடு' ஜூன் 23 முதல் 27 ஆம் தேதி வரை கோவை மாநகரில் நடைபெறுகிறது. இம்மாநாட்டின் ஒரு பகுதியாகத் தமிழ் இணைய மாநாடும் நடைபெறுகிறது.

அறிவியல் நுட்பங்களைத் தன்னகத்தே உள்வாங்கி, மாறிவரும் உலகச் சூழலுக்கேற்ப, வாக்கிலும், அச்சிலும் மட்டுமின்றி, தகவல் தொடர்பு மின்னணுச் சாதனங்களிலும், கணினி, இணைய வலைதளம் ஆகியவற்றிலும் இடம் பெற்றுள்ள மொழியாக வளர்தமிழ் விளங்குகிறது. இத்தகைய சீரிளமைத் திறமும், தகவமைவுத் திறனும், மாறுதல்களைத் தன்னகத்தே ஈர்த்துத் தேவைக்கேற்பப் பரிணாம வளர்ச்சி பெறும் தன்மையும் மொழியின் நிலைத்த தன்மைக்கு இன்றியமையாதவையாகும்.

தமிழ் இணைய மாநாடு, நமது செந்தமிழ் மொழியானது மின்னணுக் கருவிகளிலும், கணினியிலும், இணைய வலைதளங்களிலும் இன்னும் அதிக அளவில் பயன்படுத்தப்படுவதற்குத் தேவையான முன்னோடி கருத்துகளை, திட்டங்களை, நுட்பங்களை நமக்கு அளிக்கும் என்பதில் ஐயமில்லை. உலகத் தமிழ்ச் செம்மொழி மாநாட்டின் மிகச் சிறந்த அம்சங்களில் ஒன்றான 'தமிழ் இணைய மாநாடு' தொடர்பாக 'தமிழ் இணைய மாநாட்டு மலர்' வெளியிடப்படுவது மிகவும் பொருத்தமானது; பாராட்டுக்குரியது.

மாநாடு வெற்றி பெற என் நல்வாழ்த்துக்கள்.


16/6/2010
(மு.க. ஸ்டாலின்)



ஆ. இராசா
 ए. राजा
 A. RAJA

मंत्री
 संचार एवं सूचना प्रौद्योगिकी
 भारत सरकार,
 इलेक्ट्रॉनिक्स निकेतन, 6, सी.जी.ओ. कॉम्प्लेक्स,
 नई दिल्ली-110 003
 MINISTER
 COMMUNICATIONS & INFORMATION TECHNOLOGY,
 GOVERNMENT OF INDIA,
 ELECTRONICS NIKETAN, 6, C.G.O. COMPLEX,
 NEW DELHI-110 003

வாழ்த்துரை

தமிழக முதல்வர் மூதறிஞர் கலைஞர் அவர்களின் பெரு முயற்சியால் கோவையில் சூன் 23 முதல் 27 வரை நடைபெறும் உலகத்தமிழ் செம்மொழி மாநாட்டின் கூறாக நடைபெறும் தமிழ் இணைய மாநாடு வெற்றி பெற எனது வாழ்த்துக்களை தெரிவித்துக் கொள்கிறேன்.

உயர்தனிச் செம்மொழி எனும் நிலையினை தமிழ்மொழி பெற்றிட முழு முதற்காரண கர்த்தாவான முத்தமிழறிஞர் கலைஞரின் ஆட்சி காலங்களில்தான் தமிழுக்கு எல்லாத் துறைகளிலும் வளர்ச்சியும் ஏற்றமும் ஏற்படும் என்பதன் தொடர்ச்சியாக கணினித் தமிழும் கைகோர்த்து வளர்கிறது என்பதற்கு அடையாளம் இம்மாநாடு.

மத்திய அரசின் தகவல் தொழில் நுட்ப வளர்ச்சி திட்டம் (Technology Development for Indian Language Programme - TDIL) ஒன்று வகுக்கப்பட்டு, கணினி சார்ந்த தொழில்நுட்ப மொழி ஆராய்ச்சி, மென்பொருள் வடிவமைப்பு (Software Tools), பிறமொழி மாற்றுத்தொழில் நுட்பம் (Interoperability) ஆகியவை குறித்து ஆய்வுகளும், ஆக்கப்பணிகளும் நடைபெற்று வருகின்றன. அனைத்து மொழிகளுக்கான ஓர்மை தொழில்நுட்பம் (Unicode) குறித்து எல்லா மாநில அரசுகள் மற்றும் தொடர்புடைய நிபுணர்கள் / நிறுவனங்களோடு கலந்தாலோசித்து, உரிய மாற்றங்களை கொண்டுவர தேவையான நடவடிக்கைகளுக்கு மத்திய அரசு அனுமதியளித்துள்ளது, இம்மாநாட்டிற்கு உரம் சேர்க்கும் நிகழ்வாகும்.

'முன்னை பழமைக்கும் பழமையாய்'

'பின்னை புதுமைக்கும் புதுமையாய்'

விளங்கும் தமிழ்மொழி இணைய தளங்களை (website) அதிக எண்ணிக்கையில் கொண்டுள்ள மொழிகளில் ஒன்று என்பது உலகளாவிய அளவுகோளில் தமிழின் ஆளுமையை பறைசாற்றுகிறது.

தமிழ் மொழி பரிவர்த்தனைக்கான பாஷை மட்டுமல்ல; அதன் சங்கக் குறியீடும் குறிக்கோளும் உலகின் வேறெந்த மொழிகளிலும் இணையாக கருத முடியாத அளவிற்கு பெருமை கொண்டது என்ற அடிப்படையில், இணைய தளத்தின் வாயிலாகவும் தமிழ் உலகிற்கு ஆற்றுகின்ற, ஆற்றப் போகின்ற பங்களிப்பை இம்மாநாடு உறுதி செய்யும் என்று நம்புகிறேன்.

இம்மாநாட்டினை நடத்திட வழிவகை செய்த மூதறிஞர் கலைஞர் அவர்களுக்கு எனது வணக்கத்தையும் நன்றியினையும் இந்திய அரசின் தகவல் தொழில்நுட்பத் துறையின் சார்பில் தெரிவிக்க கடமைப்பட்டுள்ளேன்.

தமிழ் வாழ்க! மாநாடு வெல்க!

புதுதில்லி

தேதி: 22.5.2010

அன்புடன்,

 (ஆ. இராசா)



தமிழ் இணைய மாநாடு
அறிவியல் நகரம் அலுவலகம்,
கோளரங்கம் வளாகம், சென்னை - 600 025



மு. ஆனந்தகிருஷ்ணன்
தலைவர்

வாழ்த்துச் செய்தி

இன்று நம் செந்தமிழை 'தகவல் தொழில்நுட்பம்' தான் உலகெங்கும் பரப்பியுள்ளதென கூறலாம். புதுமையான தொழில்நுட்ப சக்தியின் பயன்கள் எல்லா மக்களுக்கும் சென்றடைவதற்கு ஏதுவாக தமிழின் கணினி பயன்பாட்டைக் குறித்தே தகவல் தொழில்நுட்பம் அமையவேண்டும். பாமர மக்கள் கணினியில் ஆங்கில அறிவு இல்லாமல் தமிழைப் பயன்படுத்தி தகவல் பெற உதவ வேண்டும். தற்போதைய பல கணினித்தமிழ் ஆராய்ச்சிகளும், முன்னேற்றங்களும் அதை நோக்கி இருக்கின்றன.



மற்ற மொழிகளைவிட தமிழ் மொழி எவ்வளவு பின் தங்கி இருக்கிறது என்பதை உணர்ந்த பிறகு தாம் இழந்த வாய்ப்புகளை பெற தமிழ்ச் சமுதாயம் விழித்திருக்கிறது. இந்தத் தமிழ் இணைய மாநாட்டிற்கு சமர்ப்பிக்கப்பட்ட பெரும்பாலான ஆய்வுக் கட்டுரைகள் இணையத் தமிழில் வளர்ந்துள்ள ஆராய்ச்சி முயற்சிகளின் முன்னேற்றத்திற்கு எடுத்துக்காட்டாக அமைந்துள்ளது. கோவை தமிழ்ச் செம்மொழி மாநாட்டில் தமிழ் இணைய மாநாட்டை சேர்க்க எடுத்துக் கொள்ளப்படும் ஊக்கமும் ஒத்துழைப்புமே இதற்குச் சான்றாகும். இதற்கு நாம் நமது முதலமைச்சர் டாக்டர் கலைஞரின் பெருவாரியான ஊக்கத்திற்கும் தமிழக அரசாங்கத்தின் இணையற்ற ஆதரவுக்கும் நன்றி தெரிவிக்க வேண்டும். உலகத் தமிழ்ச் செம்மொழி மாநாடோடு இணைந்து நம் தமிழ் இணைய மாநாடு 2010 நடைபெறுவதை எண்ணி பெருமையடைகிறேன்.

தமிழ் இணைய மாநாடுகள் பல உலகெங்கும் நடத்தப்பட்ட பிறகு பெரும்பாலான சவால்களை கணினித் தமிழில் நாம் சந்தித்துள்ளோம். உலக அளவில் தமிழ் கணினித்துவத்தையும் மற்றும் இணைய பயன்பாட்டில் உருவாகும் புதிய வளர்ச்சிகளை தொடர்ந்து கவனிக்க வேண்டிய தேவையை உணர்ந்ததனால் தமிழிற்காக உத்தமம் என்ற தகவல் தொழில்நுட்ப அமைப்பு உலக அளவில் உருவாக்கப்பட்டது. அன்றிலிருந்து அந்த அமைப்பு தமிழில் இணைய முன்னேற்றங்களில் பல இடையூறுகளையும் பிரச்சனைகளையும் தாண்டி பாராட்டுதலுக்குரிய சேவையை செய்திருக்கிறது. தமிழ்க் கணினி பயன்பாட்டில் ஏற்படும் கடும் பிரச்சனைகளை உலகளவில் சந்திக்க இந்த ஒரே அமைப்பு தான் உள்ளது.

தமிழ் இணையத்தில் பல்வேறு புதிய சாதனைகள் படைக்க ஆர்வமும் எதிர்பார்ப்புகளும் வளர்ந்து வருகின்றன. இதற்காக கூட்டுறவு மனப்பாங்கும், பெருந்தன்மைக் கொண்ட நல்லுணர்வும் இன்றியமையாதது. இம்மாநாடு அதை வலியுறுத்தும் என்று முழுமையாக நம்புகிறேன். என் மனமார்ந்த வாழ்த்துக்கள்.

மு. ஆனந்தகிருஷ்ணன்.

மு. ஆனந்தகிருஷ்ணன்
தலைவர்-தமிழ் இணைய மாநாட்டுக்குழு



வாழ்த்துச் செய்தி

உலகின் பல்வேறு நாடுகளில் பேசப்படும் ஒரு பன்னாட்டு மொழி தொன்மை வாய்ந்த, நம் மொழி, நம் செம்மொழி, நம் தமிழ் மொழியாம். இந்த சிறப்பு வாய்ந்த மொழிக்கு மேலும், சிறப்பு சேர்க்கும் வகையில் மாண்புமிகு தமிழக முதல்வர் டாக்டர் கலைஞர் அவர்களால் செம்மொழி அங்கீகாரம் பெறப்பட்டது.

இந்த செம்மொழி உலகத்தவரால் போற்றப்பட வேண்டும், பாராட்டப்பட வேண்டும். பயன்படுத்தப்பட வேண்டும் என்பதை எடுத்துச் சொல்லும் விதமாக கோவையில் 2010 சூன் 23 முதல் 27 வரை நடைபெறவுள்ள மாநாட்டின் ஒரு அங்கமாக இணைய மாநாடும் நடைபெறுகிறது.

உலகில் பல்வேறு இடங்களில் வாழும் தமிழர்கள் அரிச்சுவடி முதல் ஆராய்ச்சி கல்வி வரை தமிழ் மொழி கற்கவும், தமிழர் வரலாறு, கலை, இலக்கியம், பண்பாடு ஆகியவற்றை அறிந்து கொள்ளவும், மாண்புமிகு தமிழக முதல்வர் டாக்டர் கலைஞர் அவர்களால் தொடங்கப்பட்ட பன்னாட்டு தமிழ் பயிற்சி இணையம் வாயிலாக வகுப்புகளை நடத்தி தமிழை வளர்த்து வருகிறது.

இன்று உலகம் முழுவதிலும் பரவியுள்ள தமிழர்களால் இணையத்தின் மூலம் தமிழ் பயன்படுத்தப்பட்டு வேகமான வளர்ச்சியை கண்டுள்ளது என்பது தெள்ளத்தெளிவாகிறது.

இந்த மாநாட்டின் வாயிலாக உலகத் தமிழரை ஒருங்கிணைக்கவும், உலகில் தமிழின் நிலை அனைத்துத் துறைகளிலும் மேலும் உயரவும் பல முயற்சிகள் எடுக்கப்படுகிறது. அனைத்து நாடுகளிலும் வாழுகின்ற தமிழர்கள் ஒரே குடும்பமாக வாழ்ந்து வருவதற்கு தமிழ் இணையம் பலவகைகளிலும் துணையாக இருந்து வருவதை எண்ணி மட்டற்ற மகிழ்ச்சி அடைகிறேன்.

இந்த மாநாடு தமிழின் வெற்றி மாநாடாக, தமிழரின் வெற்றி மாநாடாக சிறப்புற அமைவதற்கு வழிவகுத்த தமிழக முதல்வர் டாக்டர் கலைஞர் அவர்களுக்கு என் வணக்கத்தையும் நன்றியினையும் தகவல் தொழில் நுட்பத் துறையின் சார்பில் தெரிவித்துக் கொள்கிறேன்.

டாக்டர் பூங்கோதை ஆலடி அருணா
தகவல் தொழில்நுட்பவியல் துறை அமைச்சர்

வாழ்த்துச் செய்தி

தமிழ் ஒரு செம்மொழி. அதற்கு வளமான பண்பாடும், இலக்கியப் பாரம்பரியமும் உண்டு. பல மொழிகள் பேசப்படும் உலகில், பொருத்தமுள்ளதாக இருப்பதை கட்டிக் காக்கவும், செழித் தோங்கி வளரவும், புலம் பெயர்ந்து வாழும் தமிழர்களுடன் தன்னைத் தொடர்புபடுத்திக் கொள்ள தொழில்நுட்பத்தின் உந்து சக்தியால் இயங்கும் இவ்வுலகில் தமிழ் மொழி பலவாறாகப் பயன்படுத்தப்பட வேண்டும். விரைவாக மாறி வரும் நமது உலகில் தகவல் தொடர்புத் தொழில்நுட்பம் ஏற்படுத்தும் தாக்கம் புரட்சிமயமானது.



தனது தோற்றத்திலிருந்து, உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம் (உத்தமம்), கணினி தொழிலாளர்கள், தமிழ்க் கல்வியாளர்கள் மற்றும் தமிழ் ஆர்வலர்களுக்காக தமிழ் இணைய மாநாடுகளை நடத்துவதில் கணிசமான அளவில் பங்களித்துள்ளது. கோயம்புத்தூரில் நடைபெறும் உலகத் தமிழ் செம்மொழி மாநாட்டுடன் 9-வது தமிழ் இணைய மாநாட்டை நடத்தும் முடிவு, புலம் பெயர்ந்த தமிழர்கள் உட்பட தமிழ் பேசும் மக்களுக்கு ஒரு வரலாறு மிக்க தருணமாகும். இது, ஒன்றுகூடிக் கொண்டாடுவதற்கும் தமிழ் மொழியை புதிய உச்சத்துக்குக் கொண்டு செல்வதற்குமான நல்ல பொருத்தமான காலம்.

நாடுகளின் தேசிய எல்லைகளுக்கு அப்பாற்பட்டு தமிழ் பேசும் மக்களிடையே உள்ள தொடர்புகளையும், ஒத்துழைப்பையும் வலுப்படுத்துவதற்கு இந்த மாநாடு பங்களிக்கும் என நான் நம்பிக்கை கொண்டுள்ளேன். கடந்த காலத்தைப் போலவே, உத்தமம், மின்னிலக்க உலகச் சூழலில் தமிழ் மொழி முன்னேற்றமடைவதற்கான முக்கிய முன்னோக்கப் பார்வையுடன் பங்களித்து வருகிறது என்பது தெள்ளத் தெளிவாகத் தெரிகிறது.

இந்தப் பொருள் நிறைந்த நிகழ்ச்சியை ஒருங்கிணைத்து நடத்துவதற்கான பாராட்டுக்குரிய முயற்சிகளுக்காக சிங்கப்பூர் அரசின் சார்பில், தமிழ் நாட்டு மாநில அரசையும் இணைய மாநாட்டின் 9-வது ஏற்பாட்டுக் குழுவையும் நான் வாழ்த்த விரும்புகிறேன். கருத்தரங்கு, ஏற்பாட்டாளர்கள், படைப்பாளர்கள், பேராளர்கள், பங்கு கொள்வோர் மற்றும் நலன் விரும்பிகள் ஆகியோருக்குப் பயனளிக்கும், வளமுட்டும் ஒன்றாக அமைய வாழ்த்துகிறேன்.

இனிய பாராட்டுகளுடன்,

திரு.எஸ்.ஈஸ்வரன்,

மூத்த துணை அமைச்சர்,
வர்த்தக தொழில் அமைச்சு, கல்வி அமைச்சு,
சிங்கப்பூர்



MENTERI SUMBER MANUSIA MALAYSIA

(Minister of Human Resources, Malaysia)

மலேசிய மனித வள அமைச்சரும், மலேசிய இந்தியர் காங்கிரஸ் தேசிய உதவித் தலைவருமான மாண்புமிகு டத்தோ டாக்டர் எஸ். சுப்பிரமணியம் அவர்களின் வாழ்த்துச் செய்தி

அன்புடையீர்,

வணக்கம். தமிழ் நாட்டின் கோவை நகரில் 2010-ம் ஆண்டு ஜூன் 23 தொடங்கி 27 வரை நிகழவிருக்கும் உலகத் தமிழ்ச் செம்மொழி மாநாட்டின் ஓர் அங்கமாக 9-வது தமிழ் இணைய மாநாடு ஏற்பாடு செய்யப்பட்டிருப்பதை முன்னிட்டு வெளியிடப்படும் சிறப்பு மலருக்கு என் வாழ்த்துச் செய்தியினை வழங்குவதில் பெருமகிழ்வடைகின்றேன்.

தமிழ் அற்புதமான ஓர் அறிவியல் மொழி. இணையத்தள பயன்பாட்டிற்கு உகந்த மொழி என்பதை இன்று நாம் நமது அன்றாட வாழ்வியல் நடைமுறைகளில் கண்கூடாகக் காணமுடிகிறது. தமிழ் இலக்கணமும், இலக்கியமும் செழுமிய வளமான மொழி என்பதால்தான் உலகில் எத்தனையோ மொழிகள் இருந்த தடம் தெரியாமல் மறைந்துபோய் விட்டாலும் தமிழ் மொழி இன்றளவும் உலகின் முதன்மையான மொழிகளுள் ஒன்றாக அதுவும் செம்மொழியாக வலம் வந்துகொண்டிருக்கிறது.

நவீன கணினி உலகிலும் தமிழ் மொழியின் தாக்கம் நமக்குப் பெருமையும்-புகழும் சேர்க்கும் உன்னதமான நிலையில் இருப்பது நமக்கு மகிழ்ச்சியளிக்கின்றது. நமது வருங்கால தலைமுறையினர்க்குத் தமிழ் மொழியின் அருமை-பெருமைகளையும், இம்மொழி அறிவியல் மொழி என்பதையும் உணர்த்தி, போற்றிக் காத்திட இது போன்ற பிரம்மாண்டமான மாநாடுகளும், கருத்துப்பரிமாற்ற அரங்குகளும் நிச்சயம் பெரும் பங்காற்றிடும் என்று உறுதியாக நம்புகிறேன். இத்தகைய நகழ்ச்சிகளை முன்னின்று நடத்தும் அனைத்து அறிஞர் பெருமக்களும், தமிழ் அறிஞர்களும் என்றென்றும் நம் போற்றுதலுக்குரியவர்கள் என்பதில் எவ்வித ஐயமுமில்லை.

9-வது உலகத் தமிழ் இணைய மாநாடு வெற்றிகரமாக நடந்தேற என் இனிய நல்வாழ்த்துக்களைத் தெரிவித்துக் கொள்வதில் மகிழ்ச்சியடைகின்றேன்.

அன்புடன்,

(டத்தோ டாக்டர் சுப்பிரமணியம்)

31 மே 2010

தமிழில் நியமங்கள்-அடிப்படையிலான தகவல் தொழினுட்பப் பாவனை

ரெஷான் தேவபுர

பிரதான செயலாற்று அலுவலர்

தகவல் தொடர்பாடல் தொழினுட்ப முகவர் நிலையம்

இலங்கை மக்கள் அனேகருக்கு, ஒரு சந்தர்ப்பம் வழங்கப்படின் தத்தமது சுய மொழிகளில் தகவல் தொடர்பாடல் தொழினுட்பத்தின் பலன்களை அனுபவிக்க விரும்பின், அவ்விருப்பை அனுகூலமாக்குவதே தகவல் தொடர்பாடல் தொழினுட்ப முகவர் நிலையத்தின், சுய மொழிகள் முயற்சியின் கொள்கையாகும்.

தகவல் தொடர்பாடல் தொழினுட்ப முகவர் நிலையம், தகவல் பெறும் வழியை மேம்படுத்தும் பொருட்டு 600 க்கும் மேலான 'நனசல' (அறிவகம்)- அதாவது சமூகங்களின் தகவல் தொடர்பாடல் தொழினுட்பத் தேவைகளைப் பூர்த்தி செய்யும் பொருட்டு பல் சேவை நிலையங்களை நாடளாவிய ரீதியில் நிறுவியுள்ளது.

விஞ்ஞானம் மற்றும் தொழினுட்ப அமைச்சு, 300 க்குக் கிட்டிய விதாதா வள நிலையங்களை நிறுவி, சிறிய மற்றும் மத்திய வியாபார நிறுவனங்களில் பொருத்தமான தொழினுட்பங்களை அறிமுகப் படுத்துவதன் மூலம் சுய தொழில் வாய்ப்பை மேம்படுத்தியுள்ளது. ஆனாலும் சிங்களம் மற்றும் தமிழ் மட்டுமே பரிச்சய முள்ள பாவனையாளர்கள், இவற்றில் அவர்களுக்குத் தேவையான உள்ளடக்கம் சுய மொழி அனுசரணையுடன் இருப்பின் மட்டுமே ஆர்வத்துடன் உபயோகிப்பர்.

தமிழ் மொழி, இலங்கையின் இரு தேசிய மொழிகளில் ஒன்றாகும். ஆதலால் தமிழில், இலங்கையர் களின் நியமங்கள்-அடிப்படையிலான தகவல் தொழினுட்பப் பாவனையை உறுதிசெய்ய வேண்டியா யிற்று. தமிழில், நியமங்கள்-அடிப்படையிலான தகவல் தொழினுட்பப் பாவனையை உறுதிசெய்ய, தகவல் தொடர்பாடல் தொழினுட்ப முகவர் நிலையமானது பொது ஆலோசனை முறைவழிகள் மூலம் தமிழ் விசைப்பலகையைத் தரப்படுத்தி, தமிழ் வரிசைக்கிரம ஒழுங்கினையும் உருவாக்கியது. மேலும் தகவல் தொடர்பாடல் தொழினுட்ப முகவர் நிலையம், இலங்கைக் கட்டளைகள் நிறுவனத்துடன் இணைந்து தகவல் பரிமாற்றத்திற்கான தமிழ் வரியுருக்குறித் தொடர்பான, இலங்கைக்கான கட்டளைகள் வரைவை SLS 1326 : 2008 யுனிகோட் பதிப்பு 5.1 க்கு இயைவாக உருவாக்கியுள்ளது.

அவசர நிலைமைகளில் எவ்வாறு செயல்படுவது, மற்றும் வழங்கப்படக்கூடிய தீர்வுகள் பற்றிய அடிப்படைத் தகவல்களை வழங்கும் முதல்தவி இணையத்தளமும் 'www.firstaider.org' மற்றும் வயது வந்தோருக்கும் இணையவருக்கும் இனவிருத்திச் சுகாதாரம் பற்றிய நம்பிக்கையான தகவல்களை வழங்கும் "ஹப்பி லைப்", 'www.happy-life.org' இணையத்தளமும் தமிழ் உள்ளடக்கத்துடன் உருவாக்கப்பட்டுள்ளது.

ஆகவே தமிழ் மொழியையும் கணினியையும் மேம்படுத்தும் 'தமிழ் இணைய மாநாடு' ஆண்டுதோறும் இடம் பெறுவது சாதகமானதே. இது நியமங்கள் -அடிப்படையிலான தகவல் தொழினுட்பத்தைக் கணினியில் உபயோகிப்பவர்களுக்கு மிக முக்கியமானதும் மிகப் பயனுடையதுமாகும். இவ்வருடம் ஆனியில் இடம்பெற விருக்கும் தமிழ் இணைய மாநாட்டின் வெற்றிக்கு இலங்கை அதன் மனமார்ந்த நல்வாழ்த்துக்களைத் தெரிவிக்கின்றது.



கே. எஸ். ஸ்ரீபதி, இ.ஆ.ப.ப.
தலைமைச் செயலாளர்



தலைமைச் செயலகம்
சென்னை-600 009

வாழ்த்துச் செய்தி

இதுகாறும் 8 இணைய மாநாடுகள் பல்வேறு நாடுகளில் நடந்துள்ளது. 9^{வது} மாநாடாக, இங்கு 'உத்தமம்' அமைப்பு மற்றும் தமிழக அரசின் தகவல் தொழில்நுட்ப துறையாலும் இணைந்து நடத்தப்படும் இந்த மாநாடு ஒரு பொருள் பொதிந்த ஒன்றாக அமையும் எனக் கருதுகிறேன். இம்மாநாடு வெற்றி பெற வாழ்த்துகிறேன்.

பள்ளி மாணவர்களுக்கான அசைகலை போட்டிகளும், கல்லூரி விக்கிபீடியா போட்டிகளில் சிறப்பாக தொடங்கப்பட்டுள்ளது குறித்து மெத்த மகிழ்ச்சி. கோவையில் 23 சூன் 2010 முதல் 27 சூன் 2010 முடிய 125 அரங்குகளில் கண்காட்சிகள் நடைபெறவிருக்கும் நிகழ்ச்சிகள் வெற்றி பெற வாழ்த்துகிறேன்.


தலைமைச் செயலாளர்,
தமிழக அரசு.



உலகத் தமிழ்ச் செம்மொழி மாநாடு

கோவை, 2010 ☆ சூன் 23-27

தலைவர் : கலைஞர் மு. கருணாநிதி, மாண்புமிகு தமிழக முதலமைச்சர்

தலைமைக் குழு

:- துணைத் தலைவர்கள் :-

பேராசிரியர் திரு. க. அன்பழகன்
மாண்புமிகு தமிழக நீதியமைச்சர்

திரு. மு.க. ஸ்டாலின்
மாண்புமிகு துணை முதலமைச்சர்

முனைவர் வா.செ. குழந்தைசாமி
துணைத் தலைவர்,
உலகத் தமிழாய்வுக் கழகம்

திரு. ஐராவதம் மகாதேவன்
தொல்லியல் அறிஞர்

:- உறுப்பினர்கள் :-

திரு. கே.எஸ். ஸ்ரீபதி இ.ஆ.ப.,
தலைமைச் செயலாளர்

திரு. கு. ஞானதேசிகன் இ.ஆ.ப.,
முதன்மைச் செயலாளர், நிதித்துறை

:- அமைப்பாளர் :-

திரு. கா. அலாவுதீன் இ.ஆ.ப.,
தனி அலுவலர்

உலகத் தமிழ்ச் செம்மொழி மாநாடு,
தலைமைச் செயலகம்,

நாமக்கல் கவிஞர் மாளிகை, சென்னை.
தொலைபேசி : +91 44 25671233

மின் அஞ்சல் : chemmozhimaanadu2010@gmail.com

ஆய்வரங்க அமைப்புக் குழு

:- தலைவர் :-

பேராசிரியர் முனைவர் கா. சிவத்தம்பி

:- இணைத்தலைவர்கள் :-

முனைவர் அவ்வை நடராசன்
முனைவர் பொன். கோதண்டராமன்

:- செயலாளர் :-

கவிஞர் கனிமொழி
மாநிலங்களவை உறுப்பினர்

:- ஒருங்கிணைப்பாளர் :-

முனைவர் M. இராசேந்திரன்
துணைவேந்தர்,

தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்
தொலைபேசி : +91 4362 227040
நகர் : +91 4362 226159

மின் அஞ்சல் : chemmozhimanadu@gmail.com
வலைதளம் : www.ulakathamizhemmozhi.org

கா. அலாவுதீன், இ.ஆ.ப.,
தனி அலுவலர்.

தலைமைச் செயலகம்,
சென்னை - 600 009.

வாழ்த்துச் செய்தி

உலகத் தமிழ்ச் செம்மொழி மாநாட்டுடன் இணைந்து தமிழ் இணைய மாநாடு, கோவையில் 2010 சூன் திங்கள் 23 முதல் 27 ஆம் நாள் வரை நடைபெறுகிறது. இந்த மாநாட்டினை உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றத்துடன் (உத்தமம்) இணைந்து நடத்துவதில் தமிழக அரசு பெரிதும் மகிழ்கிறது.

இந்தத் தமிழ் இணைய மாநாடானது கணினியின் பயன்பாட்டில் தமிழ்மொழியின் பங்கினை மேம்படுத்துவது தொடர்பாக புதிய விழிப்புணர்வை ஏற்படுத்தியுள்ளது. இதன் ஒரு பகுதியாக, கணினி மூலம் மாணவர்களின் கலைத் திறனை வெளிப்படுத்த நடவடிக்கை எடுத்து ஆறு முதல் பத்தாம் வகுப்பு பயிலும் மாணவ மாணவியர்களுக்கு கணினித் தூரிகை வரைவுப் போட்டி நடத்தப்பட்டது. மேலும், மேனிநிலைப் பள்ளி மாணவர்களுக்கு அசைநிலை வரைகலை (ANIMATION) போட்டி நடந்தப்பட்டது. மாநில அளவில் இத்தகைய போட்டிகளை நடத்துவது இதுவே முதல் முறையாகும்.

மேலும், தமிழ்நாடு அரசும், தமிழ் விக்சிப்பீடியாவும் இணைந்து கல்லூரி மாணவர்களுக்கான 'விக்சிப்பீடியா தகவல் பக்கங்கள்' என்ற போட்டியும் நடத்தப்பட்டுள்ளது. இத்தகைய முயற்சியில் அரசு நேரடியாக ஈடுபடுவதும் முதல் முறையாகும். தமிழ்த் தகவல் தொழில்நுட்பத்தின் பயன்களை மாணவச் சமுதாயத்தினரிடம் கொண்டு சேர்ப்பதில் இந்த முயற்சி முன்னுதாரணமாகும்.

இணைய மாநாட்டின் ஆய்வரங்குகள் முக்கியத்துவம் வாய்ந்தவை. உள்நாடு வெளிநாடுகளைச் சேர்ந்த வல்லுநர்கள், பேராசிரியர்கள், ஆய்வு மாணவர்கள் உள்ளிட்ட பலர் ஆய்வுக் கட்டுரைகளை அளித்துள்ளனர். இணையத்திலும் கைபேசிகளிலும்

தமிழின் பயன்பாடு குறித்துப் பல்வேறு தலைப்புகளிலும் இந்த ஆய்வுக் கட்டுரைகள் மலர்ந்துள்ளன.

தமிழ்க் கணினி தொடர்பாக விரிவானதொரு கண்காட்சியும் நடத்தப்படுகிறது. இதில் தமிழக அரசின் பல்வேறு நலத் திட்டங்கள் இணையதளத்தின் மூலம் மக்களைச் சென்றடைவது தொடர்பான விவரங்கள் செயல்விளக்கமாக கணினியில் திரையிட்டு காண்பிக்கப்பட உள்ளன. சிறந்த மென்பொருளுக்குப் பரிசும் அளிக்கப்படவுள்ளது. இவ்வாறு இம்மாநாட்டின் மூலம் புதிய தொழில்நுட்பத்தின் மூலம் தமிழின் பயன்பாட்டை விரிவுபடுத்தும் பணியில் வல்லுநர்கள் மிக்க ஆர்வத்துடன் ஈடுபட்டு வருகின்றனர்.

மாண்புமிகு தமிழக முதலமைச்சர் அவர்களின் சீர்மிகு தலைமையில் அவரது கூர்மையான வழிகாட்டுதலின்படி இந்த மாநாடு வடிவமைக்கப்பெற்றுள்ளது. இந்த 9 வது தமிழ் இணைய மாநாடு, புதிய தடம் பதிக்கும் மாநாடாக அமைந்திட என் உளம் கனிந்த நல்வாழ்த்துக்களைத் தெரிவித்துக்கொள்கிறேன். இதற்காக உழைத்த அனைவருக்கும் நன்றி தெரிவித்துக் கொள்கிறேன்.

அன்புடன்,


(கா. அரணவதீன்) 16-2-82

P.W.C. DAVIDAR, I.A.S.,
Principal Secretary to Government,
Information Technology Department.

SECRETARIAT,
Chennai - 600 009.
Phone: 2567 0783
Fax: 25670505
e-mail: secyit.tn@nic.in



வாழ்த்துச் செய்தி

இணையம் ஒரு விளக்கு. அதன் சுடர் தரும் வெளிச்சம் அறிவு.

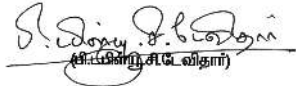
உலகிலுள்ள செல்வங்களில் உயர்ந்த செல்வம் இணையச் செல்வம்.

இதற்கு இணையான செல்வம் வேறு எதுவும் இல்லை; இதுவே ஆதாரச் செல்வம்; என்றுமே அழியாத செல்வம். பல பிறவிகளிலும் தொடர்ந்து வரப்போகிற செல்வம்; 'இணையம் என்பதே ஈடில்லாச் செல்வம்' என்ற இனிய மேற்கோள் நிகழவிருக்கிற தமிழ் இணைய மாநாடு அறிவுச் செழிப்புக்குத் தூண்டுகோலாகவும் விளங்கும் என்பதில் ஐயமில்லை.

உலகத் தமிழ்ச் செம்மொழி மாநாடு சூன் திங்கள் 23-27 வரை நடைபெறவிருப்பதை உலகிலுள்ள அனைத்துத் தமிழர்களும் அறிவர். தமிழக முதல்வரின் மைய நோக்குப் பாடல் இன்றைய தமிழ் இளைஞர்களின் எழுச்சிப் பாடலாக ஒலித்த வண்ணமாக இருப்பதைக் காணும்போது மட்டற்ற மகிழ்ச்சியடைகிறேன். தமிழ் இணைய மாநாடு சிறப்புறவும் பொலிவுறவும் நடைபெறவிருப்பதற்கு என் மகிழ்ச்சியான வாழ்த்துக்கள்

ஒவ்வொரு படி நிலையாகத் தமிழ் இணைய மாநாட்டுப் பணிகள் நடைபெறும்போது ஒருங்கிணைந்து ஆலோசனைகள் பலபெற்றுச் செவ்வனே நிறைவேற்றுவதில் மகிழ்ச்சியடைகிறேன்.

தமிழ் இணைய மாநாட்டு ஏற்பாட்டாளர்கள், படைப்பாளர்கள், பேராளர்கள் பங்கு கொள்வோர் மற்றும் நலம் விரும்பிகள் ஆகியோருக்குப் பயனளிக்கும் வளமூட்டும் கருத்துப் புதையலாக அமைய வாழ்த்து கிறேன்.


P.W.C. Davidar
தலைமைச் செயலகம்

அன்புகெழுமிய நண்பர்களுக்கு

உத்தமத்தின் தலைமைப் பொறுப்பினின்று உங்களோடு உரையாடுவதில் பெருமை கொள்கிறேன். உத்தமம் பல உத்தமர்களின் வழியில் மங்காப் புகழ்கொண்ட தமிழ் போல் தமிழிணைய வளர்ச்சியோடு வீறு நடை பயில்கிறது. முச்சங்கங்கள் கண்ட நம் தமிழ் இன்று ஒரு பெரும் வையகச் சங்கமாக இயங்குகிறது இணையம் எனும் மின்னலகில்.

இணையத்தில் இணைந்த நாம் ஒவ்வொருவரும் ஒவ்வொரு வகையில் தமிழ்க் கணினியின் வளர்ச்சியில் பங்கு கொண்டுள்ளோம் என்பதை எண்ணும் போது உள்ளம் உவகை கொள்கிறது. செம்புலப் பெயரீர் போல மின் வழி அன்புடை நெஞ்சங்கள் பல கலந்தன. தொடர்கிறது நம் நட்பும் தமிழுக்காக நம் ஒருங் கிணைந்த உழைப்பும் இம்மின்னுலகு வழியே. மின் அணுக்களினூடே அகநக நட்பில் உரை செய்த யாம் இன்று ஒன்பதாம் இணைய மாநாட்டினூடே முகநக நட்பில் திணறித் திளைக்கிறோம்.



இவ்விணைய மாநாட்டிற்கு வித்திட்டோர் பலர். இவர்கள் ஒவ்வொருவருக்கும் இவர்களின் ஒத்துழைப் பிற்காக உத்தமத்தின் சார்பாக என்னுடைய மனமார்ந்த நன்றியை முதற்கண் தெரிவித்துக் கொள்கிறேன்.

செம்மொழி மாநாட்டோடு இணைய மாநாட்டையும் இணைத்து நடத்தி நமக்குப் பெருமை அளித்த தமிழக முதல்வர் கலைஞர் அவர்களுக்கு எங்கள் மனமார்ந்த நன்றிகள். நம் முதல்வரோடு இணைந்து இணைய மாநாட்டுக்காக அயராது உழைத்த நம் பேராசிரியர் ஆனந்தகிருட்டிணன் அவர்களின் முயற்சியும் உத்தமம் தொடர்ந்து வளர அவர்களின் உள்ளக்கிடக்கையும் ஈடு இணையற்றது. உத்தமத்தினர் ஒற்றுமை நாடி முயன்றபோதெலாம் உத்தமத்திற்காக முகிலை மழைச் சாரலாக்கிய, மொட்டுகளை மலராக்கிய, இளம் பிறைதனை முழு மதியாக்கிய, பெருமை அவரையே சாரும். மின்னலகில் மண்ணுலகைப் போலவே பெரும் புயலையும் பெரு மழையையும் காணுகிறோம்! பேராசிரியர் ஆனந்தகிருட்டிணன் அவர்கள் மின்னலகிலும் நமக்காக உரைவிடம் அளித்தவரன்றோ! அவருக்கு எம் இதயம் கனிந்த நன்றிகள் என்றென்றும் உரித்தாகுக. தமிழ் இணைய மாநாட்டிற்குப் பலவகையிலும் உதவிய, தமிழ் இணைய மாநாட்டுக் குழுவில் இருந்து உதவிய அனைவருக்கும் எம் உளமார்ந்த நன்றிகள்.

ஒன்பதாவது மாநாட்டைக் காணுகிறது தமிழ் இணையம்; அதுவும் உத்தமம் பிறந்து பத்து ஆண்டுகளிலே! உலகெங்கிலும் பலப்பல இடங்களில் இணைந்தோம் நேரிடையாக. உத்தமத்தின் வளர்ச்சிக்காக நாளொரு மேனியும் பொழுதொரு வண்ணமுமாக அயராது உழைத்து, தமிழ்க் கணினியின் வளர்ச்சிக்கு வித்திட்ட அனைத்து ஆராச்சியாளர்களுக்கும் உத்தமம் தன் நன்றியைத் தெரிவித்துக்கொள்கிறது.

மேன்மேலும் வளரட்டும் நம் தமிழ்க் கணினி முயற்சிகள்! மேன்மேலும் வளரட்டும் இணையத்தில் இணைந்த நம் தமிழ் நட்பு! மேன்மேலும் வளரட்டும் உத்தமத்தின் மங்காப் புகழ்! நாம் கருத்தினிலே வேறுபடலாம். ஆனால் நம் கருத்து என்றென்றும் தமிழுக்காகவே. தமிழ்க் கணினியின் வளர்ச்சிக்காகவே. தமிழ் இணையத்தின் புகழுக்காகவே. ஒருங்கிணைவோம் என்றென்றும் இணையமெனும் மின்னலகிலிலே; தழைத்துப் பெருகிடும் அன்புடனே!

தி.ந.ச.வெங்கட ரங்கன்

தலைவர், உத்தமம்

மாநாட்டு நிகழ்ச்சிக் குழுவினின்று...

வாசு அரங்கநாதன்

பென்சில்வேனியாப் பல்கலைக் கழகம்

ஒன்பதாவது தமிழ் இணைய மாநாட்டின் மலரை இங்கு படைப்பதில் பெருமிதம் அடைகிறேன். என்றும் இல்லாத புதுப் பொலிவுடனும் வியக்கத்தக்க ஆராய்ச்சிக் கட்டுரைகளுடனும் இம்மலர் உங்கள் கையில் தவழ்கிறது. ஜெர்மனி மாநாட்டில் எண்பது கட்டுரைகள் படைக்கப்பட்டன. அதற்கும் மேலாக இம்மாநாட்டில் நூற்று முப்பது கட்டுரைகள் படைக்கப்படுகின்றன. உத்தமம் ஒரு கூட்டுக் குடும்பம் போல தமிழ்க் கணினி ஆராய்ச்சியாளர்களின் எண்ணிக்கையைப் பெருக்கிக்கொண்டே வருகிறது. இவ்வாராய்ச்சியாளர்களில் பலர் தங்களின் படைப்புகளை ஒருங்கிணைந்து படைத்துள்ளமையைக் காணும் போது தமிழ்க் கணினி மற்றும் தமிழ் இணைய ஆராய்ச்சி பல வகையிலும் ஒருமுகப்பட்டிருக்கிறது என்பது தெள்ளத்தெளிவாகிறது. இதற்கு உத்தம இயக்கத்தின் பங்கு மிக முக்கியமானது என்றால் மிகையாகாது. தொட்டனைத்தூறும் மணற் கேணி! கருத்தரங்கு காணக் காணப் பெருகிடும் தமிழிணைய வளர்ச்சி! மேலும் மேலும் உத்தம இயக்கம் வளர, மேன்மேலும் உத்தம கருத்தரங்குகள் உலகின் பல இடங்களிலும் வரும் ஆண்டுகளில் இயங்கிட உத்தம உறுப்பினர்களை வாழ்த்திடுவோம்!



இம்மலரில் கணினி மூலம் தமிழ் கற்பிப்பது என்ற தலைப்பில் இருபத்தோரு கட்டுரைகளும் அதற்கடுத்தபடியாக கணினி மொழியியல் தலைப்பில் இருபது கட்டுரைகளும் கணினி வழி தமிழ் மொழியை பகுத்தாயும் அறிவுத்திறன் கொண்ட மென்பொருட்கள் பற்றி கிட்டத்தட்ட பதின்மூன்று ஆய்வுக்கட்டுரைகளும் வந்துள்ளமை நம் தமிழ்க் கணினி ஆய்வாளர்களின் திறமையை எண்ணி வியக்க வைக்கிறது. தமிழ் இணைய மாநாட்டின் ஆரம்ப கால கட்டத்தில் தமிழ் எழுத்துரு பற்றியே ஏராளக் கட்டுரைகள் இருந்தன! கணினியில் தட்டச்சு செய்வது பற்றியே பலக் கட்டுரைகள் இருந்தன. ஆனால் இப்பொழுதோ கணினிக்குத் தமிழறிவு புகட்டும் முகத்தான் பலப்பல ஆராய்ச்சிக்கட்டுரைகள் வந்துள்ளமை தமிழ்க்கணினி உலகில் நிகழ்ந்துள்ள வியக்கத்தகு வளர்ச்சியைக் காட்டுகிறது. இயந்திரக் கைக்குத் தமிழ் எழுதும் திறனைக் கொடுத்தல், தமிழ்ச் சொற்றடர்களைக் கணினி தெள்ளத்தெளிவாகப் படித்தல், அச்சுத் தமிழை மின்தமிழுக்கு மாற்றுதல் போன்ற செயல்பாடுகள் மிக எளிதான விதத்தை இக்கருத்தரங்கின் கட்டுரைகள் பல எடுத்தியம்புகின்றன! ஒரு காலத்தில் இலக்கியங்களின் சொற்கோவையைத் தொடுக்கப் பல ஆராய்ச்சியாளர்கள் வருடக்கணக்கில் முனைந்து முனைவர் பட்டத்தோடு அவற்றை வெளியிடுவர். ஆனால் இக்காலத்திலோ ஓரிரு மணி நேரங்களில் பல்லாயிரக் கணக்கான சொற்களை மின் வடிவத்தில் உள்ள இலக்கியங்களினின்று சேகரித்து பல வடிவங்களில் பலக் கோணங்களில் நமக்குத் தரவல்ல பல தமிழ் செயலிகள் நம்மிடையே வந்துவிட்டன. இது பற்றிப் பலக் கட்டுரையாளர்கள் தமிழ்த் தரவு எனும் நிகழ்வில் தங்கள் கட்டுரைகளைப் படைக்கிறார்கள்! தமிழ்ச் சொற்களைத் தேடல், பல்லாயிரக்கணக்கான வாக்கியங்கள் உள்ள உரைகளினின்று செய்திகளைத் தொகுத்தல் போன்ற செயல்பாடுகள் மிக எளிதான ஒன்றாக ஆகிவிட்டது இக்காலக் கட்டத்தில். இது பற்றித் தமிழில் தேடுபொறி எனும் தலைப்பில் கட்டுரைகள் பல படைக்கப்படுகின்றன! கிட்டத்தட்ட தமிழ் இலக்கியங்கள் அனைத்தும் மின் வடிவத்துக்கு வந்துவிட்ட

இக்காலக்கட்டத்தில் அவற்றைப் பல கண்ணோட்டங்களில் படிக்கச் சொற்களை அலசித் தேட பலப்பல உத்திகள் நம்மிடையே வந்துவிட்டன. உத்தமத்தினரின் தமிழ்க் கணினி ஆராய்ச்சிகள் பலவும் ஆராய்ச்சியாளர்களின் ஆய்வு உத்திகளைப் பலவிதத்திலும் மாற்றிவிட்டது. இன்னமும் நாம் அச்சுப் பிரதிகளை மட்டுமே வைத்துக்கொண்டு தமிழ் ஆராய்ச்சி செய்தால் நாம் புது உலகை இன்னும் காணவில்லை என்றே பொருள் படும். மின்னணுப் புரட்சி காணும் இப்புது உலகம் நமக்குத் தமிழையும் தமிழர்களையும் தமிழ்ப் பண்பாட்டையும் நோக்கிட முப்பரிமாணக் கண்ணாடியாகத் திகழ்கிறது. உத்தமத்தால் இணைந்தோர் புதுத் தமிழிணைய உலகம் படைத்திட்டார்! சிங்கை, மலேசியா, இலங்கை, அமெரிக்கா, பிரான்சு, இங்கிலாந்து, மற்றும் கனடா போன்ற நாட்டோடு இணைந்துள்ளது நம் தமிழ் மண் இணையவுலகில்! தமிழெங்கள் இணைப்பு! மின்னிணையம் எங்கள் வல்லமை! வல்லாரும் நல்லாரும் இணைந்த உத்தமர்தம் உலகுதான் தமிழிணைய உலகம்!

தமிழ் இணைய மாநாடு 2010 நிகழ்ச்சிக் குழு உறுப்பினர்கள்

திரு. வாசு அரங்கநாதன், அமெரிக்கா | திரு. நா.தெய்வசுந்தரம், இந்தியா | திருமதி. சுபாஷினி டிரம்மல், ஜெர்மனி | திரு. கணேசன், இந்தியா | திரு. மணியம், சிங்கப்பூர் | திரு. தவரூபன், இலங்கை | திரு. ஆ. ரா. சிவகுமாரன், சிங்கப்பூர் | திரு. முத்துக்குமார் ஆறுமுகம், இந்தியா | திரு. அப்பாசாமி முருகையன், பிரான்சு | திரு. முகுந்தராஜ், ஆஸ்திரேலியா

தமிழ் இணைய மாநாட்டு 2010 உள்ளூர் ஏற்பாட்டுக்குழு உறுப்பினர்கள்

1. **தலைவர்:** பேராசிரியர் மு. ஆனந்த கிருட்டிணன், தலைவர் ஐ.ஐ.டி., கான்பூர் மற்றும் ஆலோசகர் - உத்தமம்
2. **அமைப்பாளர்:** டாக்டர் பூங்கோதை ஆலடி அருணா, மாண்புமிகு தகவல் தொழில்நுட்பத் துறை அமைச்சர்
3. **ஒருங்கிணைப்பாளர்:** திரு பி.டபிள்யூ.சி. டேவிதார், இஆப., செயலர், தகவல் தொழில்நுட்பத் துறை
4. கவிஞர். கனிமொழி, மாநிலங்களவை உறுப்பினர்
5. முனைவர் பி.ஆர். நக்கீரன், இயக்குநர், தமிழ் இணையப் பல்கலைக் கழகம்
6. திரு. மோகன், தேசிய தகவல் மையம்,மைய அரசு
7. திரு. தி.ந.ச.வெங்கடரங்கன், தலைவர், உத்தமம்
8. திரு. ஆன்டோ பீட்டர், கணித் தமிழ்ச் சங்கம்
9. திருமதி. ஸ்வரன் லதா, இயக்குநர், இந்திய மொழிகளில் தொழில்நுட்ப மேம்பாடு, மைய அரசு
10. டாக்டர். சந்தோஷ் பாபு, இஆப., மேலாண்மை இயக்குநர், தமிழ்நாடு மின்னணு நிறுவனம்

தமிழ் இணைய மாநாட்டு 2010 நிகழ்ச்சிக் குழு உறுப்பினர்கள்

1. **தலைவர்:** முனைவர். வாசு அரங்கநாதன், பென்சில்வேனியா பல்கலைக்கழகம், அமெரிக்கா
2. திரு. எஸ் மணியம், சிங்கப்பூர்
3. முனைவர். ஆ. ரா. சிவகுமாரன், நன்யாங் தொழில்நுட்பப் பல்கலைக்கழகம், சிங்கப்பூர்
4. முனைவர். ந. தெய்வசுந்தரம், சென்னைப் பல்கலைக்கழகம், இந்தியா
5. முனைவர். முத்துக்குமார் ஆறுமுகம், குமரகுரு கல்லூரி, இந்தியா
6. முனைவர். மா. கணேசன், அண்ணாமலைப் பல்கலைக்கழகம், இந்தியா
7. முனைவர். அப்பாசாமி முருகையன், Ecole Pratique des Hautes Etudes, பிரான்சு
8. திரு. தவரூபன், இலங்கை
9. திரு. முகுந்த்ராஜ், ஆஸ்திரேலியா
10. திருமதி சுபாஷினி ட்ரெம்மல், ஜெர்மனி

தமிழ் இணைய மாநாட்டு 2010 பன்னாட்டுக் குழு உறுப்பினர்கள்

1. **தலைவர்:** திரு. வா.மு.சே. கவிஅரசன், ஓகயோ, அமெரிக்கா
2. முனைவர் நாக. கணேசன், டெக்சாஸ், அமெரிக்கா
3. திரு. சிவா பிள்ளை, லண்டன், இங்கிலாந்து
4. திரு. இளந்தமிழன், மலேசியா
5. திரு. கலைமணி, சிங்கப்பூர்
6. முனைவர் மறைமலை இலக்குவனார், சென்னை
7. முனைவர் இராம. கிருட்டிணன், சென்னை
8. முனைவர் பத்ரி சேஷாத்ரி, சென்னை
9. திரு. இனிய நேரு, சென்னை.
10. திரு. தில்லைக் குமரன், கலிஃபோர்னியா, அமெரிக்கா

1

கணினி வழி தமிழ் மொழி கற்றல்

மற்றும் கற்பித்தல்



சிங்கப்பூரில் தமிழ் மொழி கற்றல் கற்பித்தலில் கணினியின் பயன்பாடு

டாக்டர் A Ra சிவகுமாரன்

இணைப் பேராசிரியர், தலைவர்,

தமிழ்மொழி பண்பாட்டுப் பிரிவு

தேசியக் கல்விக்கழகம், நன்யாங் தொழில்நுட்பப் பல்கலைக்கழகம்

சிங்கப்பூர்.

உலகில் தமிழ்மொழி பரவலாகப் பேசப்பட்டு வருகிறது. சிங்கப்பூரில் அது ஆட்சி மொழியாக இருக்கிறது; பள்ளிகளில் கற்பிக்கப்படுகிறது. இன்னும் சொல்லப்போனால் சிங்கப்பூரில் தமிழர்கள் தமிழைக் கட்டாயமாகப் படிக்க வேண்டும் என்றும் ஊக்குவிக்கப்படுகிறார்கள். தமிழ்மொழியை இனிமையாகவும் எளிமையாகவும் கற்க-கற்பிக்கக் கணினி பல வழிகளிலும் பயன்படுத்தப் படுகிறது. இப்பொழுது தமிழ்மொழியைக் கற்பிப்பதில் கணினி எந்த அளவிற்குப் பயன்படுத்தப்படுகிறது என்பதை இக்கட்டுரை மூன்று நிலைகளில் ஆராய்கிறது.

1.தமிழ்ப் பாடம் கற்பிப்பதில் கணினியின் பங்கு குறித்த கல்வி அமைச்சின் திட்டம்.

சிங்கப்பூர்க் கல்வி அமைச்சு ஒவ்வொரு பாடத்திலும் 10 விழுக்காடு கணினிப் பாடங்கள் இருக்க வேண்டும் என்று தனது முதல் முதன்மைத் திட்டத்தில் (Master Plan-1) ஒரு விதி கொண்டு வந்தது. இதனை அன்றைய கல்வி அமைச்சர் ரியர் அட்மிரல் திரு “தியோ ச்சீ ஹென்” 28 ஏப்ரல் 1997ஆம் ஆண்டு சன் டெக் சிட்டியில் நடைபெற்ற “மாஸ்டர் பிளேன் ஒன்று” அறிமுக நிகழ்ச்சியில் அறிவித்தார். அதன்படி தமிழ்ப் பாடங்களில் கணினி வழிக் கற்பிக்கப்பட வேண்டிய கூறுகள் 10% இருந்தன. பின்னர் அறிவியல் மயமாகி வருகின்ற உலகிற்கு ஏற்ற விதத்திலும் வகுப்பறைக் கலாசாரத்தில் புதுமையை ஏற்றுக்கொள்ள வேண்டிய சூழ்நிலைக்கு ஏற்பவும் கணினி வழிக் கற்பிக்கப் படவேண்டிய கூறுகள் 30 விழுக்காட்டிற்கு உயர்த்தப்பட்டன. இதனை இரண்டாவது முதன்மைத் திட்டத்தில் (Master Plan-2) அறிமுக நிகழ்ச்சியில் அன்றைய கல்வி அமைச்சர் தர்மன் சண்முகரத்தினம் 24 ஜூலை 2002இல் அறிவித்தார். தொடர்ந்து Master Plan மூன்றில் முதல் இரண்டு திட்டங்களில் கடைப்பிடித்த கூறுகளோடு ஆசிரியப் பயிற்சிக்கு முக்கியத்துவம் அளிக்கின்ற விதத்தில் இன்னும் பல கூறுகளை இணைத்து 5 ஆகஸ்ட் 2008இல் அன்றைய கல்வி அமைச்சர் திரு “இங் எங் ஹென்” அறிவித்தார். இவற்றின் மூலம் தமிழ்மொழி கற்றல் கற்பித்தல் முறையில் 30% கூறுகள் கணினி மயப்படுத்தப் பட்டன. இதனால் ஒவ்வொரு வகுப்பு அறையிலும் கணினி இணைக்கப்பட்டு அதன் வழி மாணவர்களுக்குக் கற்றல் கற்பித்தல் நடைபெற்றுக்கொண்டு வருகிறது. மாணவர்கள் தங்களுக்கென்று வலைப்பூக்கள் உருவாக்கி அதில் பாடத்தொடர்பான பயிற்சிகளைச் செய்யவும் வலைஒலிகளைக் கேட்பதன் மூலம் வகுப்பறைக்கு அப்பால் பாடம் நடைபெறவும் வாய்ப்புகள் உருவாக்கப்பட்டன.

“நம்நாடி” என்னும் இணையத்தளம் தொடக்கத்தில் தனியாகக் கல்வி அமைச்சின் தகவல் தொடர்பு தொழில் நுட்பப் பிரிவால் நிர்வகிக்கப்பட்டது. இந்த இணையத் தளம் ஏன் உருவாக்கப்பட்டது என்பதற்குரிய

காரணத்தை அவ்விணையத்தளம் இவ்வாறு கூறுகிறது. (1) தமிழ்மொழி கற்றல் கற்பித்தலை மேம்படுத்திக் கொள்ளவும் இன்பமாகக் கற்கவும் (2) பரந்துபட்ட அறிவினைப் பல வழிகளில் பெருக்கிக் கொள்ளவும் குறிப்பாகப் பண்பாடு மற்றும் நம்மைச் சுற்றிய உலக நடப்புகளை அறிந்து கொள்ளவும் (3) சுய கல்வியை ஊக்குவிக்கவும் இந்த இணையத்தளம் வழி வகுக்கும் என்று குறிப்பிடுகிறது. இவ்விணையத்தளம் இப்பொழுது கல்வி அமைச்சின் கல்வித் தொழில்நுட்பப் பிரிவால் நிர்வகிக்கப்படும் எடிமால் www.edumal.sg என்னும் இணையத்தளத்தின் கீழ்ச் செயல்படுகிறது.

இத்தளத்தில் பாடங்களைச் சிறந்த முறையில் நடத்துவதற்குக் கல்வி அமைச்சின் பாடத்திட்டப் பிரிவால் தயாரிக்கப்பட்டிருக்கும் பயிற்று கருவிகளைக் கொண்ட தமிழ்மொழி என்னும் இணையத்தளப் பிரிவும் சேர்க்கப்பட்டுள்ளது. அதாவது கல்வி அமைச்சின் பல்வேறு பிரிவுகளால் தயாரிக்கப்படும் கணினி தொடர்பான பல பயிற்று கருவிகளும் எடிமால் என்னும் இணையத்தளத்தால் ஒருங்கிணைக்கப்பட்டுச் செயல்படுகின்றன. எனவே ஆசிரியர்கள் எடிமால் தளத்திற்குச் சென்றுவிட்டால் அங்கிருந்து கல்வி அமைச்சின் எல்லா இணையத்தளத்திற்கும் சென்று விடலாம். அதற்குரிய இணைப்புகளைக் கல்வி அமைச்சு வழங்குகிறது.

கல்வி அமைச்சின் தலைமைப்பிரிவிலிருந்து தொடக்கப்பள்ளி மாணவர்களுக்காக ஒவ்வொரு மாதமும் புதிர் வினாக்கள், வண்ணம் பூசுதல் முதலான பயிற்சிகள் கொடுக்கப்படுகின்றன. மாணவர்கள் அவற்றிற்குரிய விடைகளைக் கல்வி அமைச்சுக்கு அனுப்பி வைக்கும்பொழுது தேர்ந்தெடுக்கப்பட்ட படைப்புகளுக்கு ஒவ்வொரு மாதமும் பரிசுகள் வழங்கப்படுகின்றன.

கல்வி அமைச்சின் கல்வித் தொழில் நுட்பப்பிரிவு (Educational Technology Division) தொடக்கக்காலத்தில் ஆசிரியர்களுக்கு வேண்டிய கணினி தொடர்பான பயிற்று கருவிகளைத் தயாரித்து வந்தது. தற்பொழுது அதன் இலக்கு கற்பித்தல் தொடர்பான ஆய்வை நோக்கிச் செல்கிறது. தகவல்தொடர்புத் தொழில் நுட்பத்தைக் கொண்டு கற்பித்தலில் எவ்வாறு மேம்பாடு காணலாம் என்னும் ஆய்வு நோக்கோடு “பத்து T சிகரம்” என்னும் ஒரு Portalஐ உருவாக்கி உள்ளது. அதில் தொடக்கத்தில் பத்து பள்ளிகளைச் சேர்ந்த ஆசிரியர்கள் ஒன்றுகூடிச் சீனக் கல்வியாளர்களால் பரிந்துரைக்கப்பட்ட ஒரு புதிய முறையில் கணினி வழித் தமிழ்ப்பாடங்களை நடத்தி அதன் விளைவுகள் எப்படி இருந்தன என்பதைப் பகிர்ந்து கொண்டனர். இப்பொழுது பத்துக்கு மேற்பட்ட பள்ளிகளின் ஆசிரியர்கள் இந்த ஆய்வில் ஈடுபட்டுப் புதிய முறைகளைக் கண்டாராய்ந்து வருகின்றனர்.

மிக அண்மையில் கல்வி அமைச்சு ஒவ்வொரு பள்ளிக்கூடத்திலும் பணி புரியும் ஆசிரியர்களுள் நால்வரைத் தகவல் தொடர்புத் தொழில்நுட்ப வழிகாட்டிகளாக நியமித்துள்ளது. அவர்களுக்கு மிக அண்மைக்காலம் வரை வந்துள்ள பல்வேறு தொழில்நுட்பங்களையும் கற்றல் கற்பித்தலுக்கு எவ்வாறு பயன்படுத்தலாம் என்பதைப் பயிலரங்குகளின் வழிக் கற்றுக்கொடுக்கிறது. இந்நால்வரும் அவரவர் பள்ளிகளுக்குச் சென்று ஏனைய ஆசிரியர்களுக்கு இந்நவீன முறைகளைக் கற்றுக்கொடுப்பர். இதன் வழிக் கற்பித்தலில் அரசாங்கம் கணினியில் வரும் மாற்றங்களை உடனுக்குடன் அற்றைப்படுத்திய வண்ணம் (up-to-date) உள்ளது.

எடிமால் தளத்தில் மேலும் பல வசதிகளைக் கல்வி அமைச்சு செய்துள்ளது. பாடத்தோடு தொடர்புடைய ஒளிநாடாக் காட்சிகளைக் குறிப்பாக டிஸ்கவரி சேனலிலிருந்து (Discovery Channel) வாங்கி வைத்துள்ளது. தேவைப்படுகின்ற ஆசிரியர்கள் இவற்றைத் தங்கள் பாடங்களுக்கு உடனுக்குடன் பயன்படுத்திக் கொள்ளலாம்; மேலும் மின்னிலக்கக் களஞ்சியம் என்னும் ஒன்றை உருவாக்கி அதில் ஆங்கிலச் சொற்களுக்குரிய தமிழ்ச் சொற்களையும் மேலும் பல தகவல்களையும் இணைத்துள்ளது. இவற்றின்வழிக் கற்றல் கற்பித்தல் இனிமையும் எளிமையும் மிக்கதாக அமைகிறது.

கல்வி அமைச்சின் தகவல் தொடர்புத் தொழில்நுட்பப் பிரிவில் ஒலி, ஒளி நாடாக்கள் தயாரிப்பதற்குத் தனித்தனிக் குழுவினர் உள்ளனர். இவர்கள் தமிழ்ப் பாடங்களுக்கு வேண்டிய பயிற்றுக்கருவிகளைத்

தயாரிப்பர். மேலும் இவற்றையும் பாடத்திட்டக் குழுவினர் தயாரித்திடும் பயிற்றுக் கருவிகளையும் மேலேற்றம் செய்வதற்கு ஒரு குழுவினர் உள்ளனர். இவர்கள் ஒவ்வொரு மாதமும் 18ஆம் தேதி மேலேற்றம் (Upload) செய்வர்.

கல்வி அமைச்சு தொடக்கக் காலத்திலிருந்து தயாரித்த அனைத்துவகை மின் பயிற்றுக் கருவிகளையும் e-media என்னும் இணையத்தளத்தில் வைத்துள்ளது. எனவே ஆசிரியர்கள் எல்லாக் காலக்கட்டத்திலும் தயாரிக்கப்பட்ட மின் பயிற்று கருவிகளை ஒரு குடையின்கீழ்ப் பயன்படுத்திக்கொள்ள இத்தளம் வாய்ப்பு அளிக்கிறது.

சிங்கப்பூர்த் தமிழ்க் கல்வி வரலாற்றில் கணினியைப் பயன்படுத்துவதில் நேர்ந்த மிக முக்கியமானதொரு நிகழ்வு என்றால் அது கணினியின் தமிழ் விசைப்பலகையைத் தரப்படுத்தியது ஆகும். சிங்கப்பூரில் தமிழாசிரியர்களுக்கு இடையே கணியன் அல்லது IE விசைப்பலகை, டைப்ரைட்டர் விசைப்பலகை, Romanized விசைப்பலகை, T99 விசைப்பலகை எனப் பலவகை விசைப்பலகைகள் தமிழ்மொழியை உள்ளீடு செய்வதற்குப் பயன்பாட்டில் இருந்துவந்தன. பல்வேறு விசைப்பலகைகளின் பயன்பாடு தமிழில் குறுந்தகடு (Cd Rom) தயாரிப்பதிலும் அதனை மாணவர்கள் பயன்படுத்துவதிலும் சிரமத்தை ஏற்படுத்தியது. 2003ஆம் ஆண்டிலிருந்து T99 விசைப்பலகையைப் பயன்படுத்த சிங்கப்பூர்க் கல்விஅமைச்சு ஊக்குவித்தது; இருப்பினும் இந்த விசைப்பலகையைப் பயன்படுத்துவதில் சிறுசிறு சிக்கல்கள் இருந்ததால் சிங்கப்பூர்க் கல்விஅமைச்சு மிகுந்த பரிசீலனைக்குப் பின்பு சிறு மாற்றம் செய்யப்பட்ட T99-MOE விசைப்பலகையையே அனைத்துத் தமிழாசிரியர்களும் கட்டாயம் பயன்படுத்த வேண்டும் என்னும் கொள்கையை வகுத்தது. இதனை 27 மே 2009இல் வர்த்தகத் தொழில் துறை மற்றும் கல்விக்கான மூத்த துணை அமைச்சர் திரு ஈஸ்வரன் அவர்கள் அறிவித்தார். மேலும் புதிய விசைப்பலகையை எப்படிப் பயன்படுத்துவது என்பதற்கு அனைத்துத் தமிழாசிரியர்களுக்கும் நாடு தழுவிய அளவில் பயிற்சி அளிக்கப்பட்டது. பல்வேறு விசைப்பலகைகளைக் பயன்படுத்தி வந்த ஆசிரியர்கள் படிப்படியாக T99-MOE விசைப்பலகைக்கு மாறிக்கொண்டு வருகிறார்கள் என்பது குறிப்பிடத்தக்கது. (June 2010க்குள் T99-MOEக்கு மாறிவிட வேண்டும் என்று கல்விஅமைச்சு எதிர்பார்க்கிறது). விசைப்பலகையின் தரக்கட்டுப்பாடு மாணவர்கள் இணையத்தின்மூலம் பயிற்சிகள் செய்வதற்குப் பலவிதத்திலும் உதவி வருகிறது என்றால் அது மிகையாகாது.

கல்விஅமைச்சு எதிர்காலத்தில் பள்ளிகள் எவ்வாறு அமைய வேண்டும் என்பதைச் செயல்படுத்திப் பார்க்க “எதிர்காலப்பள்ளிகள் Future Schools” என்று சில குறிப்பிட்ட பள்ளிகளைத் தேர்ந்தெடுத்து அதில் மிக நவீன தொழில் நுட்பங்களைப் பயன்படுத்திக் கற்றல் கற்பித்தலைப் பரிசோதித்து வருகின்றது. அந்த வகையில் Crescent Girls Secondary school, Jurong West Secondary school, Beacon Primary School ஆகிய பள்ளிகள் தகவல் தொடர்புத் தொழில்நுட்பக் கூறுகளைச் சோதிக்கத் தேர்ந்தெடுக்கப்பட்டன. இங்குப் பல நவீன முறையில் பாடங்கள் நடத்தப் படுகின்றன. இப்பள்ளி மாணவர்கள் ஒவ்வொருவரிடமும் ஒரு மடிக்கணினி அவசியம் இருக்க வேண்டும். பாடங்கள் கணினி மூலம் தயாரிக்கப்பட்டு மாணவர்களுக்கு அனுப்பி வைக்கப்படும். இவற்றை மாணவர்கள் வீட்டிலிருந்தவாறே படிக்க வதிகள் உண்டு.

சிங்கப்பூரில் எல்லாப் பள்ளிகளிலும் தமிழ்மொழி கற்பிக்கப்படுவதில்லை. இவ்வாறு தமிழ்மொழி கற்பிக்கப்படாத பள்ளியில் உள்ள மாணவர்கள் பக்கத்திலிருக்கும் தமிழ்மொழி நிலையங்களுக்குச் செல்வர். அத்தகைய நிலையங்களில் ஒன்று உமறுப் புலவர் தமிழ்மொழி நிலையம். இது சிங்கப்பூரின் தமிழ்மொழி வளமை நிலையமாகவும் விளங்குகிறது. இங்குத் தமிழ்மொழி கற்பிப்பதில் பல நவீன முறைகள் கடைப்பிடிக்கப்படுகின்றன. குறிப்பாக வானொலியில் பேசுவது எப்படி? அதை ஒலிபரப்புவது எப்படி? முதலான முறைகள் கற்றுக்கொடுக்கப்படுகின்றன. சில நேரங்களில் இவர்களது உரையாடல்கள் சிங்கப்பூர் வானொலியில் ஒலி பரப்பப்படும். இந்நிலையத்தில் Media Literacy Club என்னும் ஒன்று தொடங்கப்பட்டு மாணவர்களுக்கு Audio, Video, மற்றும் ஒளிபரப்பு சார்ந்த கூறுகள் கற்றுக் கொடுக்கப்படுகின்றன. இதன்

மூலம் மாணவர்கள் தமிழை அறிவதோடு தகவல்தொடர்புத் தொழில்நுட்பங்களையும் அறிந்துகொள்கிறார்கள். இங்குள்ள ஆசிரியர்கள் நிகழ்நிலைக் கற்றல் (On-line Learning) மூலம் கூடுதலான பயிற்சிகளை மாணவர்களுக்கு வழங்குகிறார்கள்.

தொடக்கப்பள்ளி மாணவர்கள் கணினி தொடர்பாக என்னென்ன திறன்களைப் பெற்றிருக்க வேண்டும் என்பதையும் கல்வி அமைச்சு வரையறுத்துள்ளது. பல்வேறு மின்னிலக்கப் பயிற்று வளங்களைக் (Digital resources and learning resources) குறிப்பிட்ட மூலங்களிலிருந்துபெற்றுப் பயன்படுத்த அறிந்திருப்பர், தேடுபொறியைப் (Search engine) பயன்படுத்த அறிந்திருப்பர், கொடுக்கப்பட்ட மீத்தொடுப்பு (Hyper Link) வழியாக இணையத்திலுள்ள வளங்களைப் பெற அறிந்திருப்பர். முக்கியச்சொற்களைக் (Keywords) கொண்டு குறிப்பிட்ட ஒன்றைப் பல்வேறு தேடுபொறிகளின் வழி தேட அறிந்திருப்பர். மாணவர்களுக்குத் தேவையான இந்தக் கணினி அறிவை ஒவ்வொரு பள்ளியும் வழங்குகிறது. மேலும் ஒவ்வொரு பள்ளியிலும் ஆசிரியர்களுக்குக் கணினி சார்ந்த உதவிகளைச் செய்ய ஓர் உதவியாள் அரசாங்கத்தால் நியமிக்கப்பட்டுள்ளார். இவர் ஆசிரியர்களுக்குக் கணினி சார்ந்த நிபுணத்துவ அறிவை வழங்குவதில் உதவியாக இருப்பார்.

2. தமிழாசிரியர்கள் தமிழ் கற்பித்தலுக்குப் பயன்படுத்தும் தொழில் நுட்பங்கள்.

கல்வி அமைச்சு தயாரித்துக் கொடுக்கும் பயிற்றுக்கருவிகளைத் தவிர்த்துத் தமிழாசிரியர்கள் தமிழ் கற்றல் கற்பித்தலில் பல்வேறு அறிவியல் சார்ந்த தொழில் நுட்பங்களைக் கையாண்டு வருகின்றனர் என்றால் அது மிகையாகாது. தொடக்கப்பள்ளி ஆசிரியர்கள் கணினியில் பலவகையான மென்பொருள்களைப் பயன்படுத்திப் பல்வேறு வகைகளில் தமிழ் கற்றல் கற்பித்தலை இனிமையாக்கிக் கொண்டு இருக்கின்றனர்.

சில பள்ளி மாணவர்கள் கட்டுரை எழுதுவதிலும் கதை சொல்வதிலும் Dream viewer Koobits என்னும் மென்பொருளைப் பயன்படுத்துகிறார்கள். மாணவர்கள் தாங்கள் பார்த்த காட்சிகளுக்குத் தக்கவாறு ஒளிப்படக் காட்சிகளையும் புகைப்படங்களையும் படல்களையும் இணைத்துக் கட்டுரை எழுதுவதற்கும் தங்கள் படைப்பாற்றலையும் கற்பனையையும் வளர்த்துக்கொள்வதற்கும் இந்த மென்பொருள்கள் பெரிய அளவில் துணைபுரிகின்றன. Koobits என்னும் இந்த மென்பொருள் Pei Tong Primary School, Ahamed Ibrahim Primary School முதலிய பள்ளி ஆசிரியர்கள் பயன்படுத்துகிறார்கள். Princess Elizabeth தொடக்கப்பள்ளியில் Inno Garden என்னும் பகுதி உருவாக்கப்பட்டு மாணவர்கள் விளையாட்டு முறையிலேயே பாடம் படிப்பதற்கு ஏற்பாடு செய்யப்பட்டுள்ளது. கணினியின் பயன்பாடு இங்கு அதிகம். மாணவர்கள் ஒரே நேரத்தில் திரையில் பார்க்கலாம் அல்லது ஆசிரியர் ஒலிபரப்பும் பயிற்சியைக் கேட்கலாம். ஆனால் எல்லாவற்றையும் ஒருங்கிணைத்துக் கட்டுப்படுத்தும் கருவி ஆசிரியரிடம் இருக்கும்.

உயர்நிலைப்பள்ளி ஆசிரியர்களான திரு ஞானசேகரன். திரு மோகன், திரு முருகன் மூவரும் இணைந்து கூகல் இணையத்தளத்தைப் பயன்படுத்தித் தமிழ் கற்பிக்கும் முயற்சியில் ஈடுபட்டிருக்கின்றனர். இதன்மூலம் ஒரு ஆசிரியர் தயாரிக்கும் பயிற்சியை மூன்று பள்ளி மாணவர்களும் பயன்படுத்திக் கொள்ளலாம். இதனால் ஆசிரியரின் தயாரிப்பு நேரம் குறைகிறது. மேலும் ஒவ்வொரு ஆசிரியரின் கற்பனைத் திறனுக்கும் புத்தாக்கத்திற்கும் ஏற்றாற்போல் பயிற்சிகள் அமைந்து மாணவர்களைக் கவரவும் வழி உண்டு. இதைவிட மூன்று பள்ளிமாணவர்களும் அந்தப் பயிற்சிக்கு எவ்வாறு விடை அளித்துள்ளனர் என்பதை மூன்று பள்ளி மாணவர்களும் பார்க்க வாய்ப்பு உண்டு. இதன்மூலம் மாணவர்கள் பல பதில்களை ஒப்பிட்டு எது சிறப்பு வாய்ந்தது ஏன் சிறப்பு வாய்ந்தது என்பதை அறிய முடிகிறது. இதனால் ஆரோக்கியமான போட்டித் தன்மை மாணவர்களிடம் ஏற்படுகிறது. இந்த முறையை இன்னும் விரிவு படுத்தும்போது பல பள்ளி மாணவர்களின் படைப்புகளையும் ஒருங்கே பார்க்க வாய்ப்பு இருக்கும்.

Multi Media Builder என்னும் மென்பொருள் கீழ்நிலைத் தொடக்கப்பள்ளி மாணவர்களுக்குச் சொற்களை அறிமுகப்படுத்தவும் பேச்சுத்தமிழைக் கற்பிக்கவும் பயன்படும் சாதனமாக உள்ளது. இலக்கியப் பாடல்காட்சிகளை இதன் வழி மாணவர்கள் மனம் கவரும் வண்ணம் கற்பிக்கலாம்.

இது மட்டுமல்லாமல் மாணவர்களிடையே வினையாடலை (interactive) ஏற்படுத்தி ஆசிரியர் ஓர் இடத்திலும் மாணவர் ஓரிடத்திலும் இருந்துகொண்டு தமிழ்ப் பயிற்சிகள் செய்வதற்கு மிக ஏதுவான முறையில் பயன்படும் Hot Potatoes, Class Marker என்னும் மென்பொருள்களைப் பயன்படுத்துகின்றனர். இம்மென்பொருளின் வழித் தெரிவு 1. விடை வினா (Multiple choice Question) பயிற்சிகள். 2. மாறிக் கிடக்கும் சொற்களைச் சரியான தொடராக்கும் பயிற்சிகள். முதலியவற்றை வழங்கிட முடியும்.

இன்றைய நிலையில் மாணவர்கள் வகுப்பில் நடந்துகொள்ளும் விதங்களைப் படம் எடுத்து அப்படங்களை வகுப்பின்முன் போட்டுக்காட்டிப் பாடம் நடத்த முடிகிறது. Digital Cameraவின் வழி இது சாத்தியமாகிறது.

Black Board என்னும் மென்பொருள் இணையம் வழிக் கற்றல் கற்பித்தலை எளிமையாக்குகிறது. ஒரே நேரத்தில் வகுப்பிலுள்ள மாணவர்கள் எல்லாரும் அவரவர் வீட்டிலிருந்த படியே கருத்துகளைப் பரிமாறிக்கொள்ளவும் ஆசிரியர் உடனுக்குடன் கருத்துகளைக் கூறவும் வழிஅமைத்துக்கொடுக்கிறது. முழுக்க முழுக்கக் கற்றல் கற்பித்தலுக்காகவே இந்த வலைதளம் இயங்குகிறது.

Web quest என்னும் இணைய வினாவிளக்கம் இன்று மிகவும் பிரபலமாகி வருகிறது. இதன்வழி மாணவர்களே கற்றுக்கொள்பவர்களாகவும் கற்பிப்பவர்களாகவும் அமைகிறார்கள். மேலும் மாணவர்களே அதைத் திருத்தி மதிப்பெண்கள் கொடுத்துக்கொள்ளவும் மாணவர்களிடத்துக் குழு உணர்வை வளர்க்கவும் குழுவாக இருந்து அறிவைப் பெறவும் இந்த இணைய வினா விளக்கமுறை பெரிதும் உதவுகிறது.

Video Editing என்னும் மென்பொருளால் பாடம் நடத்துவதற்குப் பொருத்தமானதும் தேவையானதுமான திரைப்படப் பகுதிகளைத் தனியாக வெட்டி எடுத்துக்கொண்டு அவற்றில் ஆசிரியருக்குத் தேவையான விதத்தில் ஒலி ஒளிக் கூறுகளை மாற்றி அமைத்துக்கொண்டு பாடம் நடத்தலாம்.

3. தமிழ் கற்றல் கற்பிப்பதில் வணிக ரீதியிலான பங்கு

சிங்கப்பூரில் தமிழ் கற்றல் கற்பித்தலில் வணிக ரீதியில் கணினியின் பங்கு மிக அதிகம் என்றே கூற வேண்டும். தொடக்க நிலை, உயர்நிலை மாணவர்களுக்குப் பல நிறுவனங்கள் பல மென்பொருள்களைத் தயாரித்த வண்ணம் உள்ளன. இது மட்டுமன்றித் துணைப்பாட ஆசிரியர் போன்று பாடங்களைச் சொல்லிக் கொடுப்பதோடு மாணவர்கள் சோர்வடைகின்றபோது அவர்களுக்கு உற்சாக வார்த்தைகளைக் கூறி ஊக்குவித்துப் பின்பு திருத்தியும் கொடுக்கும் மென்பொருள்களும் உள்ளன. இவற்றின் வாயிலாக மாணவர்கள் தங்களுடைய வேகத்திற்கு ஏற்பவும் புரிந்துணர்வுக்கு ஏற்பவும் தமிழ்ப் பாடங்களைக் கற்றுக்கொள்ள முடிகிறது.

உலகம் மிக வேகமாக வளர்ந்து வருகிறது. மாற்றங்கள் எல்லாத்துறையிலும் நிகழ்ந்த வண்ணம் உள்ளன. அதற்கு ஏற்பச் சிங்கப்பூரில் தமிழ் கற்றல் கற்பித்தலிலும் மாற்றங்கள் நிகழ்ந்த வண்ணம் உள்ளன. அந்த மாற்றங்களை ஏற்படுத்தக்கூடிய கருவிகளில் கணினி முன்னணியில் நிற்கின்றது என்றால் அது மிகையன்று.

மடிக்கணினியில் கன்னித்தமிழ்

ஒரு கற்றல் அனுபவம்

திரு. சம்பந்தம் மோகன் , தமிழாசிரியர்

கிரசண்ட் பெண்கள் பள்ளி, சிங்கப்பூர்

முன்னுரை

சிங்கப்பூரின் நான்கு அதிகாரத்துவ மொழிகளுள் தமிழ் மொழியும் ஒன்று. தமிழ் மொழியைச் சிங்கப்பூரில் வாழும் மொழியாகத் திகழ சிங்கப்பூர் அரசாங்கமும் சமூக அமைப்புகளும் தமிழாசிரியர்களும் பெற்றோர்களும் பெருமுயற்சி எடுத்துக் கொண்டு வருகின்றனர். தமிழாசிரியர்களுக்கும் பள்ளிகளில் முழு ஒத்துழைப்பு அளிக்கப்படுகிறது. தங்கள் பள்ளிகளில் இருக்கும் தொழில்நுட்ப வசதிகளைப் பயன்படுத்தி ஆசிரியர்களும் கற்றல் கற்பித்தலைச் சிறப்பாகச் செய்கின்றனர். அந்த வகையில் எங்கள் பள்ளி மாணவிகள் 2005-ஆம் ஆண்டு முதல் மடிக்கணினியைப் பயன்படுத்தி வருகின்றனர். நாள்தோறும் மடிக்கணினியைப் பள்ளிக்குக் கொண்டு வந்து அவற்றின் மூலம் பாடங்களைக் கற்கின்றனர். பாடபுத்தகங்களும் கரும்பலகையுமே வகுப்பறையில் இடம்பெற்ற காலம் போய் இப்போது கணினி அந்த இடத்தைப் பிடித்துவிட்டது. அந்தக் கணினியிலும்தான் எத்தனை மாற்றங்கள். அவற்றில் பயன்படுத்தும் மென்பொருள்களில்தான் எத்தனை விதம்! மலைக்க வைக்கும் அம்மாற்றங்கள் ஐம்புலனோடு ஆறாவது புலனாக கற்றலில் துணைபுரிகின்றன என்றால் அது மிகையன்று.

கணினியில் நிகழும் புதிய கண்டுபிடிப்புகளை நாம் எவ்வாறு கற்றல் கற்பித்தலில் பயன்படுத்திக் கொள்ளலாம் என்பதே நம் முன் உள்ள சவாலாகும். தொழில்நுட்பம் என்னும் அந்த மாபெரும் பூதத்தை நம் கட்டுக்குள் கொண்டுவந்து விட்டால் கற்றல் கற்பித்தலில் நம் தேவைகளை எளிதாக நிறைவேற்றிக்கொள்ள முடியும். மாணவர்கள், தொழில்நுட்பத்தைத் தங்கள் கட்டுக்குள் கொண்டு வந்து கற்றலில் உயர்நிலையை அடையவும் வழிகாட்ட முடியும்.

இன்றைய மாணவர்கள் நவீன தொழில்நுட்பத்திறனைக் கையாளுவதில் வல்லவர்கள். புதிய கண்டுபிடிப்புகளை அறிந்துகொள்வதில் ஆர்வம் உடையவர்கள். எனவே, நாம் மாணவர்களின் நிலைக்கேற்ப நம்மை மாற்றிக்கொண்டால் மாணவர்களை நம்பால் ஈர்க்க முடியும். அதிலும் நம் தாய்மொழியாம் தமிழ் கற்பித்தலில் அவர்களைக் கவர்வதற்கு இந்த மாற்றம் கட்டாயம் தேவையானதாக உள்ளது. மற்ற துறை பாடங்களுக்கு ஈடாக அல்லது ஓரளவாவது மின்னியல் பாடங்களையும் தொழில்நுட்ப வசதிகளைப் பயன்படுத்திக் கற்பித்தலை மேற்கொள்ள வேண்டியுள்ளது. தொழில்நுட்பம் வழி மின் பாடங்களை நடத்தும்போது அது மாணவர்களைக் கவர்வதோடு நெருக்கடியான நேரங்களில் அதாவது 'சார்ஸ்' மற்றும் 'சளிக்காய்ச்சல்' போன்ற தொற்றுநோய்கள் பரவும் சமயங்களில், பள்ளிகள் நடைபெறாத போது கைகொடுப்பதாகவும் அமைந்துள்ளது. அந்த வகையில், தொழில்நுட்பத்தின் ஒரு கூறாக எங்கள் பள்ளியில் 'மடிக்கணினி' கற்றல் கற்பித்தலில் பயன்படுத்தப்பட்டு வருகிறது.

நோக்கம்

சிங்கப்பூரில் கிரசண்ட் பெண்கள் பள்ளி மடிக்கணினியைப் பயன்படுத்தும் ஒரு முன்னோட்டப் பள்ளியாகும். எனவே, அக்கணினியைப் பயன்படுத்துவதில் நாங்கள் சந்தித்த சவால்கள் பற்றியும் அக்கணினி மூலம் நாங்கள் கண்ட வெற்றிகளையும் உங்களோடு பகிர்ந்துகொள்ள விரும்புகிறோம். மடிக்கணினி, புதிய மென்பொருள்கள் ஆகியன மாணவர்களின் கேட்டல், பேசுதல், வாசித்தல், எழுதுதல் ஆகிய மொழித்

திறன்கள் வளர்வதற்கு உதவின. மேலும், அவர்களின் கற்றல் ஆர்வத்தையும் தூண்டின. எனவே, எங்கள் அனுபவங்கள் நிச்சயமாக கல்வியாளர்களுக்கும் ஆசிரியப் பெருமக்களுக்கும் உதவியாக இருக்கும் என்று நம்புகிறோம்.

மடிக்கணினி தெரிவு, பள்ளி சார்ந்த காரணங்கள்:

1. கூடுதலாக உதவி தேவைப்படும் மாணவர்கள் ஆசிரியரோடும் மற்ற மாணவர்களோடும் எந்த நேரத்திலும் தொடர்புகொள்ள இயல வேண்டும்.
2. மாணவர்களின் கற்றல் எந்த நேரத்திலும் எங்கிருந்தாலும் தடைபடக்கூடாது.
3. சிங்கப்பூரில் சில நேரங்களில் திடீரென்று பரவும் தொற்றுநோய்களால் (மூச்சுத்திணறல், சளிக்காய்ச்சல்) பள்ளிகளைச் சில நாட்களுக்கு மூடவேண்டிய நிலை ஏற்படலாம். அது போன்ற சந்தர்ப்பங்களில் மாணவர்களின் பாடங்கள் பாதிக்கப்படக்கூடாது.
4. நீண்ட நாள் மருத்துவ விடுப்பில் (கை,கால் எலும்பு முறிவு போன்றவை) இருப்பவர்கள் பயன்பெற வேண்டும்.

அடிப்படை தேவைகள் :

- மாணவர்கள் அனைவருக்கும் மடிக்கணினி
- பள்ளிக்குச் சொந்தமாக இணையத்தளம் (கிரசண்ட் - முனையம் -portal)
- ஆசிரியர்கள், மாணவர்களுக்குப் பயிற்சி நடவடிக்கைகள்
- கல்வித் தேவைகளை நிறைவு செய்யும் மென்பொருள்கள்
- மென்பொருள்கள் சார்ந்த நிறுவனங்களின் ஒத்துழைப்பு

மாணவர்கள் அனைவருக்கும் மடிக்கணினி

கிரசண்ட் பெண்கள் பள்ளி, எதிர்காலப் பள்ளிகளுள் (Future Schools) ஒன்றாகச் சிங்கப்பூர் கல்வி அமைச்சால் 2007 ஆம் ஆண்டு தெரிவு செய்யப்பட்டது. அதற்கு ஒரு சில ஆண்டுகளுக்கு முன்பே பள்ளி தொலைநோக்குத் திட்டத்தோடு வகுப்பறை வசதிகள், 21 ஆம் நூற்றாண்டுச் சவால்களை எதிர்நோக்கும் பாடதிட்டங்கள், தொழில்நுட்ப வசதிகளைக் கற்றல் கற்பித்தலுக்குப் பயன்படுத்துதல் போன்ற பல நடவடிக்கைகளை மேற்கொண்டுள்ளது. தொழில்நுட்பப் பயன்பாட்டின் ஒரு கூறாக இணையம் வழி மின்-பாடங்கள் அமைத்து அவற்றைச் செயல்படுத்துவதற்கு எங்கள் பள்ளி முதல்வரும் மற்ற ஆசிரியர்களும் பெரும் முயற்சி எடுத்து வெற்றி கண்டுள்ளனர். இத்திட்டத்தைச் செயல்படுத்துவதற்கு முதலில் மாணவர்கள் அனைவருக்கும் மடிக்கணினி இருக்க வேண்டும் என்று முடிவு செய்யப்பட்டது. மேலும், அதில் எழுதும் வசதி, படக்கருவி, மற்றும் பல்வேறு மென்பொருள்களை இயக்கும் அதிக திறன் ஆகியவை இருந்தால் மிகவும் வசதியாக இருக்கும் என்பதால் நாங்கள் அந்த வகை மடிக்கணினியை மாணவர்களுக்குப் பரிந்துரை செய்தோம். அதன்படி எங்கள் திட்டம் விரிவாக பெற்றோர்களுக்கும் மாணவர்களுக்கும் விளக்கப்பட்டது. அதன்பிறகு மாணவர்கள் அனைவரும் கணினியை வாங்கினர். எனவே, எங்கள் திட்டத்தின் அடிப்படை வெற்றிக்கு மாணவர்களும் பெற்றோர்களுமே முதற்காரணம் என்று கூறலாம்.

கிரசண்ட் - முனையம் (portal)

பள்ளி இணையத்தளத்தில் நுழைவதற்கு மாணவர்களுக்கும், ஆசிரியர்களுக்கும், பெற்றோர்களுக்கும் வசதிகள் செய்து தரப்பட்டுள்ளன. அவர்கள் தங்களின் கடவுச்சொல்லைப் பயன்படுத்தி இணையப் பக்கத்திற்குச் செல்ல முடியும். ஆசிரியர் மற்றவர்களின் பக்கத்திற்கு அதாவது மாணவர், பெற்றோர் பக்கத்திற்குச் சென்று அவரவருக்குரிய தகவல்களைப் பதிவேற்றம் செய்ய முடியும். அவற்றில்

மாற்றங்களைச் செய்ய முடியும். ஆனால், மாணவர்களும் பெற்றோர்களும் அவரவருக்குரிய பக்கங்களுக்குத்தான் செல்ல முடியும்.

பள்ளியின் இணையப்பக்க மாதிரி வடிவம்

ஆசிரியர் பக்கம்	மாணவர் பக்கம்	பெற்றோர் பக்கம்
<ul style="list-style-type: none"> பள்ளி சார்ந்த நடவடிக்கைகள் மின்னஞ்சல் பாடத்திட்டம் மென்பொருள் உதவிகள் <p>மேலும் ஆசிரியர்கள் தொடர்பான தகவல்கள்</p>	<ul style="list-style-type: none"> அறிவிப்புகள் பாடங்கள் பாடவேளை பயிற்சிகள் தேர்வுத்தாள்கள் <p>மேலும் மாணவர்களுக்குரிய தகவல்கள்</p>	<ul style="list-style-type: none"> பள்ளி நடவடிக்கைகள் பாடவேளை அறிவிப்புகள் <p>மேலும் பெற்றோருக்குத் தொடர்புடைய தகவல்கள்</p>

ஆசிரியர்களுக்கும் மாணவர்களுக்கும் பயிற்சிகள்

கிரசன்ட் பெண்கள் பள்ளியில் ஆசிரியர்களுக்கு வருடத்திற்கு இரண்டு முறை தகவல் தொழில்நுட்ப பயிற்சி வகுப்புகள் நடைபெறுகின்றன. பயிற்சி வகுப்பு குறைந்தது நான்கு நாட்கள் நடைபெறும். அதில் பள்ளியில் பயன்படுத்தப்படும் மென்பொருள் குறித்த விளக்கம் அளிப்பதோடு பயிற்சியும் அளிக்கப்படும். பயிற்சிக்குப் பிறகு உதவி தேவைப்படுபவர்களுக்கு உதவி செய்ய கணினி வல்லுநர்கள் பள்ளியில் எந்நேரமும் தயாராக இருப்பர். அவர்கள் மென்பொருள் குறித்தும் கணினிப் பிரச்சனைகள் குறித்தும் தீர்வுகளை உடனுக்குடன் அளிப்பர்.

பயன்படுத்தப்படும் மென்பொருள்கள்

Lectora Professional Publishing Suite

<http://www.trivantis.com>

கற்றல் கற்பித்தல் நடவடிக்கைகளுக்குப் பல்வேறு மென்பொருள்கள் பயன்படுத்தப்படுகின்றன. அவற்றுள் 'Lectora Professional Publishing Suite' என்ற மென்பொருளைக் கொண்டு பல்லாடகக் கணினிப் பாடங்களைத் தயாரித்து அவற்றை இணையத்தளத்தின் மூலம் மாணவர்களுக்கு வழங்க முடியும். அதாவது, அதில் உருவாக்கப்பட்ட பாடங்களைக் கணினி வட்டுகளாகவும் 'எச் டி எம் எல்' கோப்புகளாகவும் மாற்றிக்கொள்ள முடியும். இதன்வழி உருவாக்கப்படும் பாடங்கள் மாணவர்களின் ஆர்வத்தைத் தூண்டும் வகையில் அமைவதோடு மாணவர்களின் சுயகற்றலுக்கும் துணை நிற்கின்றன.

VGLF (Virtual Global Learning Faculty)

http://www.heulab.com/products_funpack_series.asp

VGLF என்ற மென்பொருள் வழி மாணவர்களோடு நேரடியாகத் தொடர்பு கொள்ள முடியும். இந்த மென்பொருள் மாணவர்கள் மற்றும் ஆசிரியர்களின் கணினியில் நிறுவப்பட்டுள்ளதால் எந்த நேரத்திலும் இணைய உரையாடல் வழியாகவும் வீடியோ வழியாகவும் பாடங்களை நடத்தலாம். இணைய உரையாடல் மூலம் ஒவ்வொரு மாணவர்களும் மற்ற மாணவர்களின் கருத்துகளை அறிந்துகொள்ள முடியும். அதே சமயத்தில் மாணவர்களின் கணினியை ஆசிரியர்கள் காண்காணிக்கவும் முடியும். அவர்களுக்கு வேண்டிய தகவல்களையும் அளிக்க முடியும்.

BRAVO

<http://www.c3softworks.com/>

BRAVO என்னும் இந்த மென்பொருள் வகுப்பறை கற்றல் நடவடிக்கைகளுக்குப் பெரிதும் உதவுகின்றது. விளையாட்டு முறையில் கற்றல் நடவடிக்கைகளை மேற்கொள்ள முடியும். இந்த விளையாட்டுகள் மாணவர்களைக் கவரும் வகையில் உருவாக்கப்பட்டுள்ளன. மேலும், பயன்படுத்துவதற்கு எளிமையாகவும் அமைந்துள்ளது. இம்மென்பொருளைக் கொண்டு நடவடிக்கை மேற்கொள்ளும்போது அதில் மாணவர்கள் மிகுந்த உற்சாகத்துடன் பங்கேற்கின்றனர்.

NewsMaker

<http://www.aboutnewsmaker.com/>

NewsMaker மென்பொருள் மாணவர்களின் படைப்பாற்றலை வளர்க்க உதவுகின்றது. இதன் மூலம் மாணவர்கள் தங்களின் வாசிக்கும் திறனை வளர்த்துக்கொள்ள முடியும். திரையில் அவர்கள் எழுதிய பகுதியைப் படித்துக்கொண்டே அதைப் பதிவும் செய்ய முடியும். பிறகு அவர்கள் வாசித்த பகுதியில் பிழைகள் இருப்பின் அவற்றை நீக்கிக்கொண்டு மீண்டும் பதிவு செய்ய முடியும்.

Heulab Live Record

<http://www.heulab.com>

Heulab Live Record என்ற மென்பொருளைக் கொண்டு மாணவர்களுக்குரிய தகவல்களைப் பதிவு செய்து அளிக்க முடியும். வாசித்தல் பயிற்சி, கட்டுரைத்திருத்தம், புதிய மென்பொருளைக் கற்றுத்தருதல் போன்ற நடவடிக்கைகளையும் மேற்கொள்ள முடியும்.

Imedia

பள்ளி இணையப் பக்கத்தின் ஒரு கூறாக அமைந்துள்ள இதில் மாணவர்களின் படைப்புகள் இடம்பெறும். அப்படைப்புகள் குறித்து மாணவர்கள் தங்கள் கருத்துகளையும் பதிவு செய்ய முடியும். மாணவர்கள் கலந்துகொண்ட நிகழ்ச்சிகள் குறித்த புகைப்படங்கள், வீடியோ காட்சிகள் போன்றவற்றையும் பதிவேற்றம் செய்ய முடியும்.

பாட வடிவமைப்பும் நோக்கமும்

தகவல் தொழில் நுட்பத்தைப் பயன்படுத்தி பாடங்களை வடிவமைக்கின்றபோது அவை மாணவர்களின் ஆர்வத்தைத் தூண்டும் வகையிலும் கற்றல் விளைவுகளை மனத்தில் கொண்டும் உருவாக்கப்பட வேண்டும் என்பது எங்கள் பள்ளியின் அடிப்படை நோக்கம். எனவே, கணினிப் பாடங்களை வடிவமைக்கின்றபோது கற்றலில் அவற்றின் படிநிலைகள் எவ்வாறு இருக்கவேண்டும் என்பதை முடிவு செய்துகொள்கிறோம்.

கற்றலில் தொழில்நுட்பத்தின் படிநிலைகள் 5 E's

கற்றல் அணுகுமுறை மற்றும் சிந்தனைத் திறன் ஆகியவற்றை அடிப்படையாகக் கொண்டு கணினிப் பாடங்கள் வடிவமைக்கப்படுகின்றன.

Exchange – பரிமாற்றம் அதாவது கருத்துப் பரிமாற்றம். எடுத்துக்காட்டாக கண்கவர் இடங்கள் என்னும் தலைப்பில் அமைந்த 'நயாகரா' அருவியைப் பற்றிய பாடப் பனுவலையும் அது தொடர்பான விளக்கத்தையும் 'பவர்பாயிண்ட்' படைப்பில் விளக்குவது மற்றும் ஆசிரியர் மாணவரின் கருத்துப் பரிமாற்றத்தைக் கொண்டிருப்பது. இதன் மூலம் மாணவர்கள் பாடப் பனுவல் பகுதியைப் புரிந்துகொள்ள முடியும்.

Enrich- வளப்படுத்துதல் - கற்பித்தல் ஆர்வம் தரும் வகையில் பல்லாடக வளங்களைக் கொண்டிருத்தல். கண்கவர் இடங்கள் பற்றிய தலைப்பில் 'நயாகரா' என்னும் தலைப்பில் அமைந்த பாடப் பனுவல் மற்றும் அது தொடர்பான வீடியோ காட்சிகள், பதிவு செய்யப்பட்ட ஒலிக்கோப்புகள், படங்கள் போன்ற பல்லாடக வளங்களைக் கொண்டு கற்பித்தலை மேலும் ஆர்வம் மிகுந்ததாக ஆக்குதல். இதன் மூலம் மாணவர்களின் புரிந்துகொள்ளும் திறன் அதிகரிக்கிறது. பாடப்பொருள் சார்ந்த பல்வேறு தகவல்களைப் பெறுகின்றனர்.

Enhance - மேம்படுத்துதல்- பாடக் கருப்பொருளையொட்டி அமைந்த வேறு சில இடங்களைப்பற்றி மாணவர்கள், இணையப்பக்கங்கள் மற்றும் நூல்கள் மூலம் தகவலைத் திரட்டுதல். பின் அவற்றிலிருந்து பனுவல்களை உருவாக்குதல். (பிடித்த இடங்களைப்பற்றி எழுதுதல்) போன்ற நடவடிக்கைகளை மேற்கொள்ளுதல். மாணவர்கள் கற்றதைப் பயன்படுத்தும் முறையில் அந்நடவடிக்கைகள் அமையும்.

Extend- விரிவுபடுத்துதல் - கண்கவர் இடங்கள் என்னும் கருப்பொருளையொட்டி மாணவர்கள் பனுவல்களை உருவாக்கும் போது ஏற்படும் சந்தேகங்களைத் தீர்த்துக்கொள்ளுதல். பனுவல் பற்றிய கருத்துகளைப் பரிமாறிக்கொள்ளுதல் போன்றவற்றை இணைய உரையாடல்கள் (VGLF) மூலமாக மேற்கொள்வர். அதன்பின் பனுவலை உருவாக்குவர். இதன் மூலம் பகுத்தாய்வது மற்றும் மதிப்பிடுவது போன்ற கற்றல் திறனை வளர்த்துக் கொள்வர்.

Empower - பலப்படுத்துதல் - மாணவர்கள் உருவாக்கிய பனுவல்களைக் கொண்டு கணினி வழி படைப்புளை அளித்தல் (எ-டு) வலையொலி, வலைப்பூக்கள், பல்லாடகப் படைப்புகள் போன்றவை. இதன் மூலம் மாணவர்கள் மற்றவர்களின் படைப்புகளை மதிப்பிட அறிவர். அவர்களின் படைப்பாற்றல் திறனும் வெளிப்படும்.

முடிவுரை

இன்றைய அறிவியல் உலகில் தகவல் தொழில்நுட்பத்துறையில் நிகழ்ந்து வரும் கண்டுபிடிப்புகள் மாணவர்களிடம் பெரும் தாக்கத்தை ஏற்படுத்தி வருகின்றன என்பதை நாம் எல்லாரும் அறிவோம். கணினியும் இணையமும் கற்றல் கற்பித்தலுக்குரிய அருமையான துணைக்கருவிகள். அவற்றை நம் கட்டுப்பாட்டுக்குள் கொண்டுவந்து சிறப்பான முறையில் பயன்படுத்தினால், அவை நிச்சயமாக மாணவர்களின் முன்னேற்றத்திற்கு உதவும். இன்றைய நிலையில் மற்ற பாடங்களுக்கு நிகராகத் தமிழ்மொழிக் கற்றல் கற்பித்தல் நடவடிக்கைகளையும் கணினி மூலமாக நாம் மேற்கொள்ள முடியும் என்று உறுதியாக நம்பலாம்.

பகிர்வோம்! தமிழை வாழும் மொழியாக்குவோம்!

கல்விக்கூடங்களில் தமிழ் உபயோகத்தினைத் தரப்படுத்துதல்

Standardisation of Tamil Usage in Educational Institutions

V. Raman

B.Sc., M.A., D.F.Tech., D.L.L., DACP., Dip. in Rus., CCP., CMM., CFA., M.Phil., PhD

Chairman / Correspondent, Ananya Vidyalaya

Former Professor, Universiti Teknologi MARA, Shah Alam, Malaysia

Former Professor, Universiti Sains Malaysia, Penang, Malaysia

e-mail: raman600@gmail.com

முன்னுரை

செம்மொழியான தமிழ் உலகெங்கும் 80 மில்லியன் மக்களால் பேசப்பட்டு வருகிறது. இத்துடன் இந்தியா மட்டுமின்றி இலங்கை, சிங்கப்பூர், மலேசியா உட்பட பல நாடுகளில் கல்விக்கூடங்களில் தமிழானது பயன்பாட்டு மொழியாகப் புழங்கிவருகிறது. இந்தியாவில் குறிப்பாக தமிழகத்தில் அனைத்துக் கல்விக்கூடங்களிலும் தமிழ் மொழியின் பயன்பாடு உள்ளது. தமிழகத்திலுள்ள அனைத்துப் பல்கலைக் கழகங்களிலும் தமிழ் பற்றிய ஆய்வுகளும், தமிழில் ஆய்வுகளும் நடைபெற்றுவருகின்றன. அனைத்துப் பள்ளிகளிலும் தமிழானது, தமிழக அரசால் கட்டாயப்பாடமாக்கப்பட்டுள்ளது. சிங்கப்பூர், மலேசியாவிலும் பள்ளிகளில் தமிழ் பாடமாக உள்ளது.

ஆயினும் ஒரு கல்விக்கூடத்திலுள்ள தகவல்களையோ, தரவுகளையோ மற்றோர் இடத்தில் பயன்படுத்தவோ, கையாளவோ பெரும்பாலும் இயலுவதில்லை. இதன்காரணமாக தமிழ் ஆராய்ச்சிகள், செயல்பாடுகள், மனித உழைப்பு மற்றும் அறிவுத்திறன் ஆகியவை சிறிய வட்டத்திற்குள் சுருங்கிப்போகின்றன. பல்கலைக்கழகங்களில் உருவாக்கப்படும் ஆய்வுகள் அப்பல்கலைக்கழகத்தை விட்டு வெளியே வருவதில்லை. அனைத்துக் கல்விக்கூடங்களும் செய்த செயல்களையே, ஆய்வுகளையே, தகவல் திரட்டல்களையே திரும்பத் திரும்ப பலமுறை செய்ய வேண்டியுள்ளது. உலகமயமாதலின் (Globalisation) நன்மைகள் தமிழ்ப்பணிக்குக் கிடைப்பதில்லை. இது குறித்து இக்கட்டுரை விரிவாக ஆராய்கிறது.

கல்விக்கூடங்களும் தமிழ் உபயோகமும்:

கல்விக்கூடங்களைப் பொறுத்தவரை இந்தியாவில் தமிழகத்தில் ஆட்சி மொழியாக உள்ள தமிழ் அனைத்துக் கல்விக்கூடங்களிலும் பயன்பாட்டில் உள்ளது. மேலும் தமிழகத்திலுள்ள அனைத்து பல்கலைக் கழகங்களிலும் ஆய்வுப் பணிகளும் தமிழில் நடைபெற்று வருகின்றன. வரும் கல்வியாண்டிலிருந்து (2010-11) தமிழகத்தில் பொறியியல் படிப்புகள் தமிழில் அறிமுகப்படுத்தப்படுகின்றன. சிங்கப்பூர், மலேசியாவிலும் தமிழ் பள்ளியில் பயன்பாட்டில் இருப்பதை அனைவரும் அறிவர். இலங்கையில் தமிழ் தேசிய மொழியாகவும், கல்விக்கூடங்களில் பயன்பாட்டு மொழியாகவும் உள்ளது. இதுதவிர உலகெங்கும் உள்ள பல்வேறு பல்கலைக் கழகங்களில் தமிழ்ப் பயன்பாடு பரவலாக இருப்பதைக் காண முடிகிறது.

இவ்வாறு உலகெங்கும் தமிழ் கல்விக்கூடங்களில் பயன்பாட்டில் இருந்து வந்தபோதும் கவல்களையோ, தரவுகளையோ மற்றோர் இடத்தில் பயன்படுத்தவோ, கையாளவோ இயலுவதில்லை. மேலும் எந்த ஒரு கல்விக்கூடமோ, பல்கலைக்கழகமோ எந்தவிதத்திலும் தமிழ்ப் பயன்பாட்டால் ஒன்றிணையவில்லை.

இதற்கு முக்கிய காரணிகளாக உள்ள பிரச்சனைகளாவன குறியீட்டுப் பிரச்சனையும், தரவு கையாள்தல் பிரச்சனையும் ஆகும்.

தமிழை உலகெங்கும் பரவலாகப் பயன்படுத்தி வந்த போதிலும் ஆரம்ப காலத்திலிருந்தே தரப்படுத்துதல் (Standardisation) சரியான முறையில் கையாளப்படவில்லை. இதன் விளைவாக உலகெங்குமுள்ள தமிழர்கள் ஏராளமான தமிழ் மென்பொருட்களை அவரவர் விருப்பப்படி சொந்த குறியாக்க முறையில் தயாரித்துடன் விநியோகிக்கவும் செய்தனர். இதன்விளைவாக நூற்றுக்கணக்கான குறியீட்டு முறைகள் பழக்கத்தில் வரத்தொடங்கின. இதனால் கல்விக்கூடங்களில் தரவுக் கையாள்தலில் மிகப்பல பிரச்சனைகள் உருவாகத் தொடங்கின. இவற்றை நாம் ஆய்வதற்கு முன்பாக குறியாக்க முறைபற்றியும் அதன் விளைவுகளைப் பற்றியும் ஆராய்வோம்.

கணிப்பொறியில் குறியாக்க (Encoding) முறை :

உலகில் பல மொழிகளிலும் கணினி பயன்படுத்தப்பட்டு வருகிறது. எந்தவொரு மொழியாக இருந்தாலும் கணினியில் பயன்படுத்தும் போது ஒவ்வொரு எழுத்துக்கும் ஓர் எண் ஒதுக்கப்படும். சான்றாக, 'A' என்ற எழுத்துக்கு 01000001 என்ற எண்ணும், 'B' எழுத்துக்கு 01000010 என்ற எண்ணும் பயன்படுத்தப்படுகிறது. இதற்கு ஆஸ்கி முறை (ASCII - American standard code for Information Interchange) என்று பெயர். இதேபோல் ஒவ்வொரு மொழியில் உள்ள எழுத்துருவிற்கும் ஒவ்வொரு எண் ஒதுக்கப்படும். எண் அமைப்பினைக் கொண்டே கணினி இயங்குகிறது. இப்படிக் கணினியில் ஒவ்வொரு எண்ணையும் நிர்ணயிக்கும் குறியீட்டு முறைக்கு குறியாக்கம் (Encoding) என்று பெயர். இவ்வாறு குறியாக்கம் செய்வதில் உலக மொழியாக உள்ள ஆங்கிலம் பாதி இடத்தைப் பிடித்துக் கொண்டது.

அதாவது, எழுத்துருவிற்கு எண்ணாக ஒதுக்கப்படும் கணினியின் இடமானது 16x16 என்ற அளவில் 256 கீற்றுக்களாக (glyph) அமைகிறது. இதில் 128 கீற்றுக்களில் ஆங்கிலமொழி குறியாக்கம் செய்யப் பட்டுள்ளது. அதிலும் எழுத்துக்கள், எண்கள், நிறுத்தற்குறிகள் மற்றும் கணிதக் குறியீடுகள் என 94 எழுத்துக்கள் குறியாக்கம் செய்யப் பட்டுள்ளது. மீதமுள்ள 128 கீற்றுக்களில் மற்ற உலக மொழிகளுக்கு இடம் ஒதுக்கப்பட்டன.

தமிழ் எழுத்துரு குறியாக்கத்தின் தொடக்கம் :

இதன் அடிப்படையில் கணினியின் பயன்பாட்டில் தமிழானது தொடக்கத்தில் தமிழ் தட்டச்சுக் கருவியைத் தழுவி 'பாமினி' என்கிற எழுத்துரு அறிமுகமானது. இது ஆங்கிலக் குறியாக்கத்தில் அமைந்த எழுத்துருக்களுக்குப் பதிலாகத் தமிழை உட்புகுத்தியது. இதுபோன்றே பலரும் சொந்தத் தயாரிப்பில் பலவகை எழுத்துருக்களை உருவாக்கினர். ஆனால், காலப்போக்கில் இணையத்தின் வரவால் பல்வேறு நடைமுறைச் சிக்கல்கள் இதில் உருவாகின. இக்காலத்தில் வேறுசில குறியாக்க மென்பொருள்களும் பயன்பாட்டிற்கு வந்தன. அதனால் இணையத்தில் செய்தி பரிமாற்றத்தில் பலவகைச் சிக்கல்கள் ஏற்பட்டன.

தகுதரம் (திஸ்கி) (TSCII)

இந்த ஏற்பாட்டில், இணையத்தின் வரவு புதிய நடைமுறைச்சிக்கல்களை உருவாக்கியது. இதே காலப் பகுதியில் வேறு பல நியமங்களும் உருவாகத் தொடங்கின. இதனால் கோப்புக்களைப் (File) பரிமாறுவதில் சிக்கல்கள் ஏற்பட்டன. தவிர்த்து, பல தரவுத் தளங்களில் (Database) ஒரு எழுத்துருவை மாத்திரமே ஏற்றுக் கொள்வதால் தமிழையும் ஆங்கிலத்தையும் ஒன்று சேர்க் இயலாமல் போனது. மேலும் இவ்வகையில் தமிழ் - ஆங்கிலம் என இரு மொழிகளிலும் ஆவணம் ஒன்றைத் தொகுப்பது கடினமானது. இந்நிலையில் பல

மென்பொருள் தயாரிப்பாளர்கள், தமிழ் அறிஞர்கள், தமிழைக் கணினியில் பயன்படுத்துவோர், தமிழ் ஆர்வலர்கள் மற்றும் பலரின் முயற்சியால் புதிய தமிழ் நியமக் குறியாக்க முறை உருவாக்கப்பட்டது.

இம்முறையென்பது, “கணினி சார் தேவைகளுக்குத் தமிழ் எழுத்துக்களைப் பயன்படுத்துவதற்கென உருவாக்கப்பட்ட 8 பிட் அடிப்படையில் அமைந்த, தமிழ் - ஆங்கிலம் என இரு மொழிகளைக் கையாளத்தக்க ஒரு எழுத்துக் குறிமுறை நியமமாகும். எனவே இவற்றைக் கருத்திற் கொண்டு தகவற் பரிமாற்றத்திற்கான தமிழ் நியமக் குறியீட்டு முறை (அதாவது, Tamil Standard Code for Information Interchange [TSCII]) உருவாகியது. தமிழில் திஸ்கி எனவும், தகுதரம் எனவும் (தமிழ் குறியீட்டுத் தராதரம்) வழங்கப்படும். இதில் முதல் 0-127 எழுத்துகள் தகவல் பரிமாற்றத்திற்கான அமெரிக்க முறையை (American Standard Code for Information Interchange [ASCII]) ஒத்தது. மிகுதியான 128-155ல் தமிழ் எழுத்துகள் நிரப்பப் பட்டன. விண்டோஸ் 3.1, 95, 98, Me ஆகிய பதிப்புக்களில் TSCII அதிகம் பயன்படுத்தப்பட்டது. இதுவே முதன்முதலில் உலகம் தழுவிய இணைய உரையாடல் மூலம் தரப்படுத்தப்பட்டு உருவாக்கப்பட்ட குறிமுறை நியமமாகும்” என்று தமிழ் விக்கிப்பீடியா உரைக்கிறது.

அதுவரை நிலவிவந்த பல்வேறு சிக்கல்களுக்கு, தமிழ் எழுத்துக் குறியாக்க வரலாற்றில் ஒருமொழி (ASCII) குறியாக்கத்தால் ஏற்பட்ட வெற்றிடத்தை நிறைவு செய்ய உருவாகிய அடுத்தகட்ட வளர்ச்சியே திஸ்கி. கணினியில் ஆஸ்கி குறியாக்கத்திற்கான இடம் போக மீதமுள்ள (129 முதல் 256 வரையுள்ள கீற்றுக்களில்) இடத்தில் தமிழ் எழுத்துக்களைப் பிரதியீடு செய்தலே திஸ்கி குறியாக்கத்தின் அடிப்படை. ஆங்கில எழுத்துகளின் குறியாக்கத்தில் தமிழைப் பிரதியீடு செய்த முறையினும் இது முன்னேற்றம் உடையதாக, இருமொழிப் பயன்பாட்டிற்கு எளிதாக இருந்தது.

TAM மற்றும் TAB குறியீட்டு முறை :

கடந்த 1999ம் ஆண்டு அரசால் ஏற்றுக் கொள்ளப்பட்டு வெளியிடப்பட்ட தமிழ் 99 விசைப்பலகை அச்சமுறையில் தமிழ் எழுத்துருக்கள் TAM என்றும் TAB என்றும் இருவகையாகப் பிரிக்கப்பட்டன. TAM என்பது Tamil Monolingual. TAB என்பது Tamil Bi-lingual. TAM என்பது முழு வடிவிலான தமிழ் எழுத்துருக்களைக் கொண்டதாகும். இம்முறையில் தட்டச்சு செய்யப்படும் எழுத்துருக்கள் முழு வடிவில் இருக்கும். சான்றாக, ‘நிலா’ என்று தட்டச்சு செய்தால் ந, ி, ல, ா என்று தனித்தனி எழுத்துக்களாக இல்லாமல் நி, லா என முழு எழுத்துக்களாக இருக்கும். அழகுணர்ச்சிக்காக முழு எழுத்துக் களாக இக்குறியீட்டு முறை உருவாக்கப்பட்டது.

TAB முறையில் தமிழ் எழுத்துருக்களைக் குறியாக்கம் செய்யக் கிடைத்த 128 கீற்றுக்களில் 247 எழுத்துக்களைப் பொருத்த முடியாத காரணத்தால் எழுத்துக்களைக் கூறிட்டுப் பொருத்த வேண்டியதாயிற்று. ‘கொடு அழகு மொழிச் சொல்’ என்னும் வாக்கில் உள்ள கொ,டு,அ,ழ,கு,மொ,ழி,ச்,சொ,ல் என்னும் பத்து எழுத்துக்களைத் இக்குறியாக்கத்தில் |,க,டு,அ,ழ,கு,|,ம,ழ,டி,ச்,|,ச,|,ல் எனப் பதினேழு கீற்றுக்களாகக் குறியாக்கம் செய்தனர். இம்முறைக் குறியாக்கம் ஒரு குறையாகவேக் கருதப்படுகிறது.

ஒருங்குறி கூட்டமைப்பு (Unicode Consortium) :

ஆங்கிலம் தவிர உலகளாவிய மொழிகளில் ஏற்பட்ட இக்குறியாக்கப் பிரச்சனையைப் போக்கவும், உலகில் உள்ள அனைத்து மொழிகளையும் ஒன்றிணைக்கவும் ஒருங்குறி கூட்டமைப்பு (Unicode Consortium) என்ற அமைப்பு உருவாக்கப்பட்டது. இலாப நோக்கற்ற நிலையில் மிகப்பெரிய கணினி நிறுவனங்களும், உலக நாடுகள் பலவும், கணினி ஆர்வலர் பலரும் உறுப்பினர்களாக உள்ள இவ்வமைப்பு உலக மொழிகள் அனைத்தையும் ஒரு குடையின் கீழ் கொண்டுவரவே நிறுவப்பட்டது.

இக்கூட்டமைப்பில் உறுப்பினர்களாக உள்ளவர்கள் அவரவர் மொழியை ஒருங்குறி மூலம் கணினியில் எப்படிக் கொண்டுவருவது என்றும், அதில் ஏற்படும் சிக்கல்களுக்கு எவ்வாறு தீர்வு காண்பது என்றும் ஆராய்ந்து திட்ட அறிக்கைகளை இவ்வமைப்பிடம் வழங்குவர். அதனை ஆராய்ந்து ஒவ்வொரு மொழிக்கும்

எவ்வளவு இடம் ஒதுக்குவது என இவ்வமைப்பு முடிவு செய்யும். இவ்வமைப்பில் இந்திய அரசும், தமிழக அரசும் உறுப்பினர்களாக உள்ளன.

ஒருங்குறி குறியாக்க முறை (Unicode Encoding) :

ஒருங்குறி குறியாக்க முறை (Universal Coding என்றும் கூறுவர்) என்பது உலகளாவிய குறியாக்க முறை. இதில் உலக மொழிகள் அனைத்திற்கும் இடம் உண்டு. இது 32 பிட் திட்டம். இதில் 65000 கீற்றுக்களில் பல்வேறு மொழிகளுக்கும் இடம் ஒதுக்கப்படுகிறது. சைனீஸ் போன்ற சில மொழிகள் தவிர ஒவ்வொரு மொழிக்கும் 128 பிட்கள் ஒதுக்கப்படும். இவ்வாறு இந்திய மொழிகளுக்கு ஒதுக்கப்பட்ட குறியாக்க முறை ISCII (Indian standard code for Information Interchange) என்று அழைக்கப்பட்டது. ஆனால், இது தேவநாகரி முறையில் குறியாக்கம் செய்யப்பட்டதால் தமிழ் மொழிக்குக் குறைந்த இடங்கள் ஒதுக்கப்பட்டது. இதனால் தமிழ் பயன்பாட்டில் சில சிக்கல்கள் ஏற்பட்டன. என்றாலும் ஒருங்குறி, நமது அசிரத்தையினால் இஸ்கி சார்ந்த குறியாக்க முறையை ஏற்றுக்கொண்டுவிட்டது என்பார்கள்.

ஒருங்குறி தமிழ் குறியாக்கமும் எழுத்துரு பயன்பாடும் :

தற்போது ஒருங்குறியில் தமிழ் மொழிக்கும் 8 பிட் அடிப்படையில் குறியாக்க முறை வரையறை செய்யப்பட்டுள்ளது. இதனால் நம்மிடையே இருமொழிப் பயன்பாட்டிற்கென உள்ள TAB TSCII முறைகளை விட ஒருங்குறியில் பன்மொழி உள்ளடக்கிய எழுத்துருவைப் பயன்படுத்த வழி ஏற்படுகிறது. மேலும், TAM, TAB, TSCII போன்ற குறியாக்க முறையில் குறியீடுகள் வேற்று மொழி எழுத்துருக்களைக் கொண்டதாக இருக்கும். குறியாக்க முறையிலும் ஒரு குறியாக்க முறை 'அ' வை 140 ஆவது இடத்தில் புகுத்தி இருந்தால், வேறொரு குறியாக்க முறை 'ன' வைப் புகுத்தியிருக்கும். ஒருங்குறியில் மேற்கண்ட இடர்ப்பாடு களையப்படுவதுடன், பிறமொழிக் குறியாக்கத்தில் தமிழைப் பிரதியீடு செய்த முறையும் மாற்றம் பெற்று தமிழ் மொழி எழுத்துருக்கள் உட்புகுத்தப்படுகிறது.

ஒருங்குறியில் எண் 2944 முதல் எண் 3071 (U+DB80 - U+OBFF) வரை தமிழுக்காக இடம் ஒதுக்கப்பட்டுள்ளது. அந்நிலையில், எந்தமொழியைச் சேர்ந்தவராயினும் 2949 என்ற எண்ணை ஒருங்குறியில் எழுதினால் அது தமிழ் 'அ' வாகவே கணினியில் வெளிப்படும். மேற்கூறியவாறு உலகம் முழுவதும் ஒரே குறியாக்கம் பயன்பட்டால் செய்திப் பரிமாற்றத்தில் குழப்பமேற்படாது.

தமிழ் அனைத்து எழுத்துக்கள் குறியீட்டு முறை (TACE - Tamil All Character Encoding)

இருப்பினும் இன்றைய சூழ்நிலையில் தற்போதைய ஒருங்குறியை விடுத்து 'தமிழ் அனைத்து எழுத்துக்கள் குறியீட்டு முறை' (TACE - Tamil All Character Encoding) யையே செந்தரமாக்க வேண்டும் என்ற குரல் பலத்து ஒலிக்கிறது. இவ்வகையில் 313 வழக்கத்திலுள்ள எழுத்துருக்களும் மற்ற தமிழ் எண்கள், பழைய எழுத்துக்கள் ஆகியவைகளும் குறியீட்டு முறையில் சேர்க்கப்படும். இவற்றின் சாதக பாதக அம்சங்கள் விரிவாக அலசப்பட்டு வருகின்றன.

ஆக குறியீட்டு முறை தரவுப் படுத்தலின் வழி அனைத்துக் கல்விக்கூடங்களும் ஒரே குறியீட்டு முறைக்கு மாறிவிடும். இதனால் எழுத்துருப் பிரச்சனைகள் அறவே நீக்கப்படும். அனைவரும் தமிழை தமிழாகவே எந்த ஒரு குறிப்பிட்ட மென்பொருள் துணையின்றி படிக்க, எழுத, கையாள இயலும். கூடிய விரைவில் தமிழக அரசு ஒருங்குறியை அதிகாரப்பூர்வக் குறியீடாக அறிவிக்கப்படவுள்ளதாக அறித்துள்ளது. இது குறியீட்டுப் பிரச்சனைக்கான தீர்வாக விளங்கும் என்று கருதப்படுவது இங்கு குறிப்பிடத்தக்கது.

தரவு கையாளுதல் பிரச்சனைகள்.

தரவு கையாளுதலிலும் கல்விக்கூடங்கள் பல்வேறு சிக்கல்களைக் கொண்டுள்ளன. கல்விக்கூடங்களிடையே ஒருங்கிணைப்பானது ஏறக்குறைய அறவே இல்லை என்று கூறலாம். இதனால் ஒரு கல்விக்கூடத்தில் இருக்கும் தரவுகள் மற்ற கல்விக் கூடத்தில் கிடைப்பதில்லை. இதனால் ஆய்வுகள், தமிழ்ப் பணிகள், நூல்

வகைப்படுத்துதல், நூல் கணினிப்படுத்துதல் போன்ற அத்தியாவசியப்பணிகளை ஒவ்வொருக் கல்விக்கூடங்களும் திரும்பத் திரும்பச் செய்ய வேண்டியுள்ளது. இதனால் ஏராளமான மனித உழைப்பு, நேரம், பொருள் ஆகியவை விரயமாகின்றன. மேலும் கல்விக்கூடங்களில் உள்ள தரவுகள் பரவலாக மக்களைச் சென்று அடையாமல் அக்கல்விக்கூடங்களிலேயே முடங்கி விடுகின்றன.

தரவுக் கையாளுதல் பிரச்சனைக்கான தீர்வுகள்:

தரவுக் கையாளுதலைப் பொறுத்தவரை தீர்வுகள் சற்று சிக்கலாகவே இருப்பினும் அவை செய்யக் கூடியவையே. முதலில் அனைத்துக் கல்விக்கூடங்களும் வலைத் தொடர்பின் மூலம் ஒன்றிணைக்கப்படுதல் வேண்டும். அதுதவிர பொதுவான தரவுகள் தொடர்பான பணிகள் பகிர்ந்தளிக்கப்படுவதன்வழி விரைவில் அப்பணிகளை முடிப்பதுதான் எளிதில் கையாளவும் முடியும். எடுத்துக் காட்டாக, தமிழ் இலக்கியங்களை மின் நூலாக மாற்றும் பணியை ஒவ்வொரு பல்கலைக் கழகங்களும் தனித்தனியே செய்து தங்களின் வலைத் தளங்களில் ஏற்றி வருகின்றன. இதற்கு பதிலாக பல்கலைக் கழகங்களுக்கு பொறுப்புகள் பகிர்ந்தளிக்கப்பட்டு அவை மின் நூலாக மாற்றப்பட்டு பொது மின் நூலத்தில் சேர்க்கப்படுதல் மிகுந்த பயனளிக்கும்.

அதுதவிர தரவுத் தொகுப்புத் தொகுதிகள் உருவாக்கப்பட்டு அவைகளுக்காக இணைய தளங்களும் உருவாக்கப்பட்டு பொதுவாக்கப்படுதல் வேண்டும். எடுத்துக்காட்டாக ஆய்வுப் பணிகள், மேற்படிப்பு ஆய்வுக்கட்டுரைகள் ஆகியவைகளுக்கென ஒரு இணைய தளம் உருவாக்கப்பட்டு அனைத்துப் பல்கலைக் கழகங்களிலுமுள்ள ஆய்வு பற்றிய விவரங்களும், கட்டுரைகளும் அவ்விணைய தளத்தில் தொகுக்கப் படுதல் வேண்டும். இதனால் செய்த ஆய்வையே திரும்பச் செய்தல், ஆய்வுத் திருட்டு ஆகியவை தவிர்க்கப்படுவதுடன், எத்தகைய ஆய்வுகள் பரவலாக மேற்கொள்ளப்பட்டு வருகின்றன என்ற தகவல்களும் கிடைக்க வழிபிறக்கும்.

அனைத்துக் கல்விக் கூடங்களுக்குமான பாடநூல்கள், கேள்வித்தாள்கள், அவை தொடர்பான பிற விவரங்களை அனைவரும் எளிதில் கையாள 'தரவு கையாளுதல் தர விதிகள்' உருவாக்கப்படுவதன்வழி ஒரே மாதிரியான தரவுக் கையாளுதலை அனைத்துக் கல்விக் கூடங்களும் பயன்படுத்த இயலும்.

தொகுப்பாகப் பார்ப்பின் கீழ்க்கண்ட வழிமுறைகள் கல்விக்கூடங்களில் தமிழ் உபயோகத்தை தரப்படுத்த வழிகோலும்.

1. குறியீட்டு முறை தரப்படுத்துதல்
2. எழுத்துரு தரப்படுத்துதல்
3. பொது வலைத்தொடர்பு முறை ஏற்படுத்துதல்
4. தரவுத் தொகுப்புத் தொகுதிகள் உருவாக்கம் மற்றும் அவற்றிற்கான இணைய தளங்களை ஏற்படுத்துதல்
5. தரவு கையாளுதல் தர விதிகள் உருவாக்குதல்

எனவே தமிழ் உபயோகத்தினைத் தரப்படுத்துதல் மிகமிக அவசியமாகிறது. இக்கட்டுரையானது கல்விக்கூடங்களில் தமிழ் உபயோகத்தினை தரப்படுத்தும் வழிமுறைகள், அவற்றிலுள்ள சிக்கல்கள் மற்றும் தீர்வினை மேற்கண்டவாறு விரிவாக ஆய்கிறது. ஆயினும் உபயோகத்தின் போது ஏற்படும் பிரச்சனைகளைக் குறித்து மேலும் விரிவாக ஆராய வேண்டியுள்ளது. இது குறித்து மற்ற நாடுகளில் உள்ள கல்விக்கூடங்களில் ஏற்படும் பிரச்சனைகளைப் பற்றியும் அறிய வேண்டியுள்ளது. இவ்வாறு ஆராய்ந்து தீர்வுகள் ஏற்படுத்தின் அதன் மூலம் கல்விக்கூடங்களில் தமிழ் உபயோகத்தினைத் தரப்படுத்துதலுக்கான

வழிகிடைக்கும். இதன் மூலம் அனைத்துக் கல்விக்கூடங்களும் உலகளாவிய நிலையில் இணைக்கப்பட்டு அது தமிழுக்கும், தமிழருக்கும் மாபெரும் பேறாக அமையும் என்பது திண்ணம்.

மேற்கோள்கள்:

1. வெ.இராமன், “உயர்கல்வியில் ஒருங்குறி குறியீட்டு முறை செயலாக்குவதின் அவசியம்” சென்னைப் பல்கலைக்கழக பன்னாட்டுக் கருத்தரங்கு (2010), சென்னை, இந்தியா
2. V.Raman, “Standardising Tamil language encoding method; Analysis of technical issues” APAN International conference (2001), Malaysia.
3. V.Raman, “Communications and Tamil Language” Communication Study and the Human Sciences - A Transdisciplinary Colloquium (2003), Malaysia.
4. தமிழ் விக்கிபீடியா இணைய தளம் (www.ta.wikipedia.com)
5. ஒருங்குறி இணையதளம் (www.unicode.org)
6. K.Kalyanasundaram.Ph.D. Tamil Unicode FAQ. எழில்நிலா இணைய தளத்திலிருந்து பதிவிறக்கம் செய்யப்பட்டது.
7. எஸ்.ரெங்கராஜன் (சுஜாதா). தமிழ்க் கணினி : சில சிந்தனைகள். தமிழ் இணையம் 2003 மாநாட்டுக் கட்டுரைகள்.
8. முத்து நெடுமாறன். தமிழுக்கு சொந்த வீடு. எழில்நிலா இணைய தளத்திலிருந்து பதிவிறக்கம் செய்யப்பட்டது.

Teaching and Resource Building in Teacher Education

Dr Seetha Lakshmi

Associate Professor

Asian Languages & Cultures

National Institute of Education, Singapore

Abstract

This paper talks about the experience of teaching of Tamil language and learning through IT in pre service course training at the National Institute of Education, Singapore. Teachers are undergoing their training on educational history of Singapore, educational psychology and teaching their first and second curriculum studies with the content subjects and practicum at the pre service training. While they are going to be the teachers of 21st century learners, it is essential to equip themselves with the necessary and relevant professional skills. Ida Fajar Priyanto(2007) stated about the production IT based teaching resources for the development of teachers.

Here, instead of learning students' learning and teachers' teaching approaches, they were taught to use, facilitate with information technology and to produce resources for their students and other students. This kind of resource building providing cognitive, social and emotional constructivism based engagement and focus on a common goal i.e. developing the Tamil students in Singapore. The resources were prepared by the writing lesson but can be customized by the teachers for their teaching of other skills in Tamil class. The resources building was based on task based approach, web-quest approach, group investigation approach and multimodal approach. Although the trainees were encouraged to focus on student based learning package they also provided guidelines for the teachers to use it effectively in their class. Here, developing and equipping young students to be the frequent users of the Autonomous Technology-Assisted Language Learning (ATALL) for their understanding and learning of the second language i.e. Tamil (http://en.wikibooks.org/wiki/Autonomous_Technology-Assisted_Language_Learning). This way of learning provided the facilitation to the student while he/she learns on his/her own pace in the mode of student based learning with the communication tools for eg. Computer, sound based media and the content of their subject. A questionnaire was used to collate the trainees constructive comments as they are told to use their and their peers' resources during their teaching Practicum at various Primary Schools for 10 weeks in this January Semester, 2010.

The article will share the full picture of this process at the conference.

Introduction

Today, information technology has kept the world under its control and has made us all dance to its tunes. We should not forget one thing here. The man, who invented it, when he runs after it, takes on the role of a parent when he strives for its love and comes under its influences to ensure its growth. How do we bring information technology into the education system, and specifically, into Tamil education? Let's look at some the thoughts on this.

As the usage of information technology is prevalent in English, similar efforts are also expended in Tamil lessons. This is commendable. Nevertheless, many questions arise when we look at research sources based on the extent of a student's involvement on information technology in language learning and usage as well as the extent of information technology usage in classroom conversations. I use computers for my teachers and educators in Tamil lessons, my students carry out these activities weekly and some of that can be listed down. However, it sets one thinking on whether a student converses with a computer, or whether students converse with one another to complete assignments or whether a teacher converses with a student to make learning Tamil enjoyable or encourage a student to continue learning with the computer by telling him that he is doing well. This is because our students are well versed with the computer. Today, a five-year-old child knows how to set up a face book account. The child also knows how to change to a new password if he forgets the old one. He knows how to create his own blog. He knows how to chat online with others. However, with these skills, we cannot claim proudly that our child knows everything. The child should know his cultural, language and national boundaries well and does not cross these limitations with information technology. He should know how to protect his mind and body even when he uses this medium. He should learn how to communicate face-to-face even after conversing with the text messages on the hand phone and computer. He should know how to make a stand with his own identity in an environment with people of multi-nationalities and multi-languages. Knowing all these aspects is important; it will be futile without this knowledge.

It is here in information technology that they say that research in Tamil education has not progressed as much as we have. Although the Tamil teachers encourage the students with their love on the Tamil language and passion on their job and strong beliefs on their students' development, still there are areas to improve. At some time, based on the fluent use of written Tamil during the classroom conversation, a teacher should not mention that his/her students are using Spoken Tamil in a confident manner. After teacher training, teachers need to develop them further to excel in their job and equip themselves with the pedagogical approaches. Later they could go into undertaking studies on their students' learning and teach new things. Yes, we need more research studies in Tamil and need to progress in many areas as they can convey much information for future generations.

Let's compile information on usage on information technology in classrooms:

From 1988, NIE has been holding workshops and conferences related to information technology. Today, Tamil teachers are well-versed in both computers and the English language. But, they must also be well-versed in the Tamil language. This is important. Computers can be used to teach conversations in Tamil. These technological talents are necessary in the 21st century(). It is good to ponder if we can create a good author with the usage of a computer.

We should not use the power point software merely as a tool. We should use it as a thinking guide. Lessons should be designed to suit their ages and tap on their experiences and should also allow for their views to be aired. (Gopinathan S., 2000). He further stated that it appears that we do not spend much attention on language pedagogy that understands students' needs and prepares them for global changes and Singapore's long-term visions. In a bilingual Singapore, many have set out to learn about their culture and identity in a language that they are well-versed in. In Singapore's context, this language turns out to be English. (Gopinathan S., 2000). Here, in our today's Singapore context, a number of computer-

related issues have been resolved. Today, Tamil language should become easily conversant in the classrooms. Are Tamil teachers using computers in the classrooms?

When we focus on teachers, there are challenges that they face. There are some major ones. They are:

- Lack of time
- Inability to spend time solely on teaching as they have other duties
- The absence of appropriate atmosphere or time to read more books to enhance their teaching standards
- Inability to interact with teachers from other schools during holidays as their personal lives do not allow for such luxuries (65 per cent of Tamil teachers in Singapore are women who spend much time with their families)

If we hope to do something for teachers, we can guide teachers to make our generations intelligent. Because, through education, we should prepare our students to live in the real world. Then, with their knowledge and enriched characters, they should make changes to the world and guide the generations to come. The skills of this 21st century are mandatory for this to be made possible. These are information technological skills, discussion tactics and teamwork. We need to tailor our curriculum plans to nurture, appreciate as well as to think of one issue in many angles at any one time. Here, the NIE's new approach on using PB works and is a boon for many of our Tamil teachers' resource production based hopes.

PB Works

This section highlights various ways of using free, web-based software PB works which allow the users to increase their resources and provide passive permission to other educators to use them with proper acknowledgement and vice versa.

PB Works is software which allows a community to interact and develop further through net. The following are some of the selected features of PB works:

- It is a online community based collaborative and controlled website
- It allows everybody to take ownership and feel empowered
- It provides recognition and encourages competitiveness
- Helps to develop 21st century survival skills (collaborative, soft skills including IT skills and critical thinking skills)
- It has complete access control
- Easy adaptation
- Encouraging learning in a fun way
- Effective audit trailing which assures the citizens feel more comfortable and copyright
- Can do banking in the copyrighted materials without any fear
- "The best part is that my students have taken ownership of their wiki. Their writing has improved because they have the ultimate audience- Victoria C, High school Teacher,"(<http://pbworks.com/content/edu+resources>)
- Free training
- Sharing feelings and critical comments about the content and new initiatives of trainees

According to its website source, currently PBworks “hosts over 300,000 educational workspaces, and has helped transform teaching and learning for millions of students, parents and teachers. (<http://pbworks.com/content/edu+overview>)”



There is a sample website on how to use PB works for teaching and learning in a useful way (<http://pbworks.com/content/casestudies-academic>) and websites on the current use of other educational institutions. <http://pbworks.com/content/casestudy-northcolonie>

Preparation process:

Currently NIE is celebrating its 60th anniversary and it has made history through these 60 years of development and moving higher with more theory based practice oriented teacher training and research initiatives. Recently, at our NIE, we have produced a collaborative report titled, TE 21 and it shares its six key recommendations on refreshing, updating and strengthening NIE’s model of teacher education. This covers the initial Teacher Preparation to all the way up to Leadership Training. In other words, it provides training and equips the future teachers with the expected and relevant 21st century skills and following are the main recommendations:

1. Focus on refreshed **values, skills and knowledge** as necessary pre-requisites for the 21st century teacher.
2. Define a set of professional benchmarks as a framework for developing **teacher competencies**.
3. Strengthen the **theory-practice nexus** through mentorship and reflective teaching, among other things.
4. Extend teachers' **pedagogical repertoire** of instructional strategies, modelled after best practices, to keep abreast of changing content.
5. Develop a high level of **assessment literacy** in response to changing pedagogies, so teachers can effectively evaluate student outcomes.
6. Enhance pathways and opportunities for **professional development** to make teaching a profession of choice. (<http://singteach.nie.edu.sg/issue-22-janfeb-2010-/206-growing-a-21st-century-teaching-force.html>)

Based on these 6 recommendations, something has been tried in my Tamil pedagogical modules during the last semester and this semester. For the Diploma in Education year 1 and year 2 trainees (28+15=43), I have used them for their pedagogical development. I have explained about the challenges faced by the

Tamil teachers and within a minority community, we discussed about their needs too. One of them is getting or producing suitable resources and uses them effectively in the class. Hence this process in teacher training is very handy and timely.

First year and second year students pursuing a Diploma in education course at the National Institute of education, have been told about the resource building and in 2009 they were provided with the training to use the PB which is a website which creates a controlled website. Here, our officer at the Centre for Information Technology in Education provided the basic training to the student teachers. Based on it they have used it on this calendar year and during January semester. The planning are given here:

Preparation Process

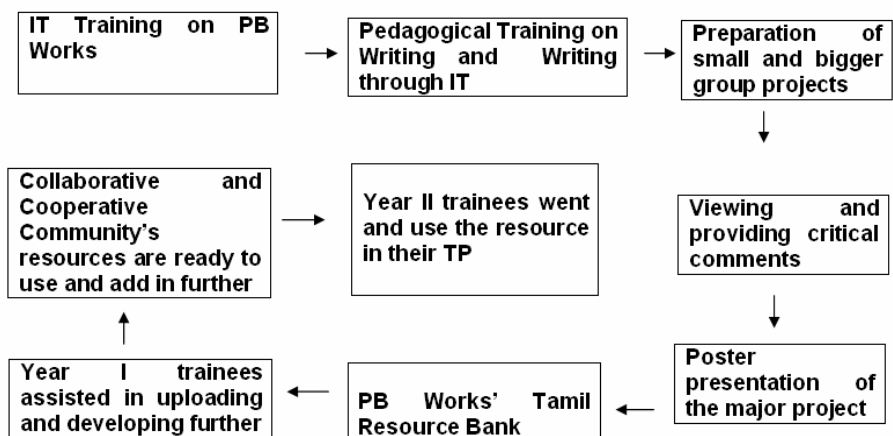
Teaching writing through IT for Dip Ed II class

Although this process involved two groups of trainees, the active involvement is limited to Dip Ed II trainees. I have taught Use of Tamil in Teaching Module (DLT200) for 12 hours within 6 weeks of time span and the following are the main topics taught :

- How to incorporate Information Technology in the Tamil Classrooms(Primary Level)
- How to develop the 21st century soft skills among the students
- Explanation of the PB Works and the Resource Building Initiative
- Creating IT based self learning package(2 for Continual assessment and one for the major project)
- Sharing of projects on writing and information technology.

Overall, the Learning skills, psychology, pedagogy, proper usage of computer, usage of multimedia, suitable use of spoken and written language were the important areas to be looked into assessing the self learning package. Here, the assessment of produced works was done at peer and lecturer level. During the assessment round, they presented their individual and group projects in their classes first. They were told to hold discussions in forum styles at that time.

PB Works in Tamil Resource Bank(Process)



The discussion forum covered areas on which sections were good and why, which sections could be better and why. These improvements were then made and uploaded on the Internet. They created websites with their names and uploaded them. They could then view their classmates' projects as necessary and learn from them. What happened here was that many uploaded projects were done out of their own interests too. This is a commendable initiative. The following picture shows the frame work of the process:

Current Development

Here, currently more than 50 items were uploaded in the PBworks website. In uploading the content, the year 1 trainees were very happy to upload their seniors' works with the lecturers' close supervision. In the meantime, they have created their own pages within their folders and started to chat within themselves.

They have shared the following comments:

- Very useful
- Motivating them to create more resources and upload
- Indirectly happy to view that their classmates and seniors visited their pages
- Continuously encouraged to talk to their PBworks mates

Diploma Year 2 trainees:

- It is good that I can place my works including other lessons
- It is a readymade resource building
- There is no copyright issues
- It is safe and controlled hence I need not worry
- During our TP time, we have used them
- Although not always, I used for my certain lessons
- Now I have a place to banking my existing resources
- I will do the same to my students and motivate them to use IT for learning Tamil in a fun way

As a professor I feel that this is a good initiative. As there is no copyright issue or no outsider interruption, I am too motivated to parking all my existing projects, power points in this website

For self - centered learning, student based learning, interactive learning, constructive and collaborative learning, this kind of resource building sites are useful and they provide lifelong learning and understanding within their digital natives' community and outside community. Here, we could witness that many trainees have been placed their personal life related useful sharings and memories. In the sharing on the 'pokkisham', I myself have learned many useful things about my trainees, their intellectual thoughts and of course their heart and mind. It is a learning curve for many others too. I found that these kinds of features are very useful in this process. This PB works have the following key elements:

Constructing knowledge: Constructivism is building the knowledge the way we see the world. The new experience will become knowledge and scaffolded with the existing knowledge. Through this the human beings are enhancing their understanding the developing their cognitive potential to be an active citizen of his/her community. Here, the trainee teachers themselves construct their knowledge on content, pedagogy and real life related authenticity. At the same time, their future students will also learn the same skills and if possible some new knowledge from this resource.

Interactivity: <http://www.clomedia.com/features/2008/December/2464/index.php?pt=a&aid=2464&start=9644&page=4>(accessed) Interactivity is a feature which is crucial to the success of the social network and it has to be done by proper and planned process. To develop the students as confident speakers and also confident writers, this process will be a good platform and it will be a role model for the Tamil Diaspora outside India and Srilanka.

Social networking: Social networking is vital for making necessary connection between the trainees. The inter and intra social networking enhances the responsibility and the integrity. Here in this project, this is happening and the schools are accepting as the trainee teachers used the products in their Teaching practicum at nearly 28 schools. So, in June another 15 will be trying this pattern of social network and it will become broader and deeper in the following years. As a result, our trainees will create their own PB works for their individual classes and they too invite their all Tamil classes and Tamil teachers to create a social network. At the end, 28X28X6= 4710 groups will be spanned and form a bigger network as *aalamaram* in Tamil. This is a short term result for us and indeed it will be a bigger and stronger pool for more resources. Yes, this will be a stronger and steady force for many predictable changes in the Tamil students' thought, cognitive, psychological process.

Conclusion

To say that there are no resources is one thing. And to make good use of available resources is another thing. Showing others the usage of the resources and giving permission for this usage is another. At bringing together these aspects is a noble act. By doing this, all our trainee teachers have this facilities in their schools. It is notable that through this, there are many ways they can urge their students to create similar projects and discuss in depth about projects already presented.

References

1. Gopinathan, S., 2000. Keynote address at the Tamil language Seminar on Teaching Tamil Language in a Fun and interesting Way. Singapore: National Institute of Education.
2. Ida Fajar Priyanto. 2007. Developing IT-based teaching materials to enhance information skills and knowledge awareness among students. World Library and Information Congress: 73rd IFLA General Conference and council. 19-23 August 2007, Durban, South Africa
3. <http://www.ifla.org/iv/ifla73/index.htm>
4. <http://pbworks.com/content/edu+resources> (accessed on 17.04.2010)
5. http://en.wikibooks.org/wiki/Autonomous_Technology-Assisted_Language_Learning. (accessed on 16.02.2010)
6. <http://archive.ifla.org/IV/ifla73/papers/133-Priyanto-en.pdf.e-essay>. (accessed on 18.04.2010)
7. <http://www.clomedia.com/features/2008/December/2464/index.php?pt=a&aid=2464&start=9644&page=4>(accessed on 17 04 2010)

8. <http://www.clomedia.com/features/2008/December/2464/index.php?pt=a&aid=2464&start=9644&page=4>(accessed on 17.04.2010)
9. <http://pbworks.com/content/edu+overview>
10. <http://singteach.nie.edu.sg/issue-22-janfeb-2010-/206-growing-a-21st-century-teaching-force.html>(accessed on 17.04.2010)
11. <http://pbworks.com/content/casestudies-academic>. (accessed on 14.04.2010).
12. Sugiarto Joesoef, 2009. School Leadership Challenges Towards Learning for 21st Century. Keynote Address at the 1st Regional Conference on Educational Leadership and Management on Globalization: Current Trends in Educational Leadership and Management. 10-12 November 2009. (unpublished essay).

Use of Technology in Running a Tamil School in USA

Ilango Meyyappan
Principal, California Tamil Academy,
Fremont Branch, California, USA

Introduction

California Tamil Academy's (CTA) primary activity is to teach Tamil to the children and young adults living in America. As an extension to teaching Tamil, CTA also supports developing cultural awareness activities such as music, dance, drama and any art form based on Tamil language and Tamil culture. CTA's mission is to develop a love for learning Tamil that will last a lifetime and develop communication skills in Tamil so that the Tamils in America can appreciate family values and feel the binding with extended family members living in India as well as the US. CTA creates a forum for our youngsters to meet, learn, share and practice our culture and values whereby providing an identity to them. CTA nurtures and preserves our cultural identity and heritage to maintain essential family values and develop self-esteem and pride around the identity by participating in Tamil school and community activities. CTA helps children feel, internalize and be self-assured about the value of Tamil and remove the question of "Why should I learn Tamil?"

CTA objectives

CTA aims to develop and improve speaking, reading and writing Tamil. Teaching Tamil is much more than just teaching letters and sounds. CTA ensures that the kids achieve small successes every week and that parents are aware of their children's progress in reading, writing and talking in Tamil. CTA ensures that knowledge passes on to successive generations. CTA believes that the kids learn Tamil culture along with the Tamil language. CTA creates a kid friendly environment for students to learn Tamil and has a "Commitment to Excellence" attitude and philosophy in teaching Tamil.

CTA History

CTA was started in 1998 with 13 students. Today, CTA has around 1800 students. CTA was started in Cupertino, California. Now, it is being run in 6 branches in Cupertino, Fremont, San Jose, Folsom, Foster City and Pleasanton. On top of this, CTA has affiliated schools in Novato, Atlanta, Seattle and Phoenix. CTA maintains a student teacher ratio of 8:1. We have around 250 teachers that teach Tamil every Sunday on a purely voluntary basis. With an ever increasing Tamil population in the US, CTA hopes to set an example as a role model school and is sure that Tamils living in other states in the US will eventually follow CTA model, syllabus and curriculum.











Use of technology

For such a massive operation that is run by volunteers who have other regular full time jobs, the use of technology is very essential in ensuring a smooth operation of the school. We will see how technology is

used in the various aspects of running this school and the benefits seen by both the parents and the teacher body by using technology.

User profiles

Every person that is in some way associated with the school can set up a user profile that has the basic information of the individual like his name, address and contact phone numbers. Each person is assigned a “role” by the management and the system. Based on the “role”, each person has certain accessibility restrictions and privilege to a certain kind of information. For example, a parent has only a “parent” role that limits access only to information regarding his or her own child and not any other child. A “teacher” role provides access only to that particular class and not any other class. A “principal” role provides access to all students, teachers, classes and reports only for that particular branch and no other branch. This ensures privacy of personal information and data.

User Role Information	
View	User Role Name
	Admin
	Teacher
	Volunteer
	Parent
	Reg Officer
	Principal
	Attendance Officer
	Book Officer
	TVU Officer
	Treasurer

View My Profile	
User Name	Demo Principal
User Email	ctawebdemo@gmail.com
User Alternate Email	-
Gender	Male
Street	Street
City	City
State	CA - California
Zip/Postal Code	30092
User Home Phone	123-456-7890

School profile

CTA uses the regular public schools, private schools and college campuses in the US for classrooms. To create a school like atmosphere and bring seriousness into teaching and learning Tamil, CTA uses regular schools by renting them for 4 hours every Sunday morning. Information about the school like its name, address and location are stored in the system. The classrooms that are being used and the grades and sections for which these rooms are assigned are also stored in the system. The teacher information is also stored. This is so that management is aware of the school and the classes that are being run in each of these rooms.

Student registration

View Payment Details				
SchoolSite	Grade	Section	Student Name	This Payment
Fremont	Grade V1	A	Akila Nalappan	200.00
Total Amount : \$				200.00
Payment Date			9/6/2009	
Payment Mode			Credit Card	
Bank Name			-	
Check No			-	
Amount			200.00	

View Student Details for : Shrijan			
Student ID	2003	School Site	Demo School
First Name	Shrijan	Last Name	Swaminathan
Grade	Preschool 1	Section	A
Gender	Male	Date of Birth	11/18/2003
Birth Place	San Jose, CA	Father Name	Swaminathan Krishnamoorthy
Mother Name	Madhu Swaminathan	Street	502 Deerpoint Ter
City	Fremont	State	CA - California
Zip/Postal Code	94536	Home Phone Number	510-790-8313
Mobile Number 1	408-341-5215	Mobile Number 2	408-341-5216
Parent Email ID	swaminathan_k@hotmail.com		
Doctor Name	Dr. Sujatha Tipness	Doctor Phone Number	510-490-1222
Dentist Name	Precious Dental Care	Dentist Phone Number	510-790-7900
EC1 Name	Vijaay Dooraiswamy	EC2 Name	Subramanian Krishnamoorthy
EC1 Relation	Uncle	EC2 Relation	Uncle
EC1 Street	637 Amberfiled ter	EC2 Street	1166 Nilda Ave
EC1 City	fremont	EC2 City	Mountain View

CTA stopped accepting manual registration using a printed application form and switched to paperless and online registration. From 2009 onwards, all registrations were done online and the payments were also made online. When a student is registered, important information about him is stored in our system and management can access it anytime to contact a student. The school has all the emergency contact information about all registered students so that the student can be helped in case of an emergency.

Student performance

Teachers maintain an online logbook to monitor and track the performance of the students on a weekly basis. Attendance, classroom participation, reading skills, homework submission are all assessed on a weekly basis and marks for those are entered in the online logbook every week. Monthly test scores are entered every month. At the end of a particular term and at the end of a school year, a report card is automatically generated. The teacher can publish a report card at the end of every term so that the parents can know what scores their children have earned.

LogBook Information				
Student Name	Working Date	Homework	Reading / Conversation	Participation
Adithya Kumar	01/17/10	10.0	10.0	10.0
Arvind Kandasamy	01/17/10	10.0	10.0	10.0
Deiva Dham	01/17/10	10.0	10.0	10.0
Kirthi Ravi	01/17/10	10.0	10.0	10.0
Nivethana Narayanan	01/17/10	10.0	10.0	10.0
Sanjana Murthy	01/17/10	10.0	10.0	10.0
SATHYAPRIYA SARAVANAN	01/17/10	10.0	10.0	10.0
Shreyas Srinivasan	01/17/10	10.0	10.0	10.0
Siddharth Raju	01/17/10	10.0	10.0	10.0

Attendance Details for School : 'Fremont' Grade : 'Grade IV' Section : 'A' Date : '01/17/2010'			
Grade	Section	Student Name	Status
Grade IV	A	AdithyaJayachandran	P
		DhanyaDuraivelu	P
		KathirIlango	P
		KryptokaranGanesan	P
		MonishaRavindar	P
		NehalVeerabose	P
		NivethaSivaprakasam	P
		VidyalakshmiMohanraj	P
		VijayRaveendran	P
No of Record(s) :9			

 California Tamil Academy								
Student Progress Report Card for Academic Year 2009 - 2010								
School: ' Fremont ' Teacher Name: ' Ganesh Lakshmanan ' Grade: ' Basic III ' Section: ' A								
Term	Student Name	Attendance	Home Work	Reading / Conversation	Participation	Monthly Test	Total	Grade
I	Adithya Kumar	9.09	10.00	5.00	5.00	68.25	97.34	A
I	Arvind Kandasamy	10.00	10.00	5.00	5.00	68.25	98.25	A
I	Deiva Dham	10.00	9.09	5.00	5.00	64.75	93.84	A
I	Kirthi Ravi	10.00	10.00	5.00	5.00	66.50	96.50	A
I	Nivethana Narayanan	9.09	9.55	5.00	5.00	66.50	95.14	A
I	Sanjana Murthy	10.00	9.36	5.00	5.00	66.50	95.86	A
I	SATHYAPRIYA SARAVANAN	10.00	10.00	5.00	5.00	66.50	96.50	A
I	Shreyas Srinivasan	9.09	10.00	5.00	5.00	66.50	95.59	A
I	Siddharth Raju	10.00	9.55	5.00	5.00	68.25	97.80	A

Communication

Teachers use web based system for communicating with the parents. The system allows the teachers to selectively choose certain parents or mass email all the parents. Teachers can also post homework details and share documents with the parents. The principal also uses the web based system to send parent and teacher weekly updates that contain information about the school, syllabus, testing schedule, model question papers and make general announcements. The model question papers are uploaded before the monthly tests. The actual question papers are also uploaded and the schools download them and print them out. Since CTA has many branches and affiliation schools, the use of technology is absolutely essential for administering tests.

Multimedia in Syllabus

There is a separate syllabus committee that takes inputs from the teachers, parents, students and other Tamil educators and sets up a syllabus that is tailored to kids growing up in a foreign land where their main form of instruction and learning is in English. CTA uses a lot of multimedia to make learning visual and intuitive for kids growing up in the US. CTA developed and printed its own set of books for reading, writing and exercise for preschool I, II and Basic I, II and III. For grade I to VII, we use the textbooks recognized and used by the government of Tamil Nadu. For almost all grades, we have the textbooks on CDs. We also use other audio and video CDs which have songs, rhymes, poems and pictorial illustrations of letters, words and sentences that make learning easier and interesting.

High school credit program

CTA recently got approval for teaching Tamil to high school students towards satisfying the foreign language requirement. With this approval, Tamil students that are in high school can take Tamil as a language and satisfy their foreign language requirement. We use the same system for administering our high school credit program teachers and students. So far, our students have been taking either Spanish or French predominantly for foreign language requirement, now they are happy that they can take their own mother tongue and earn credits for that. Only few languages have earned this qualification with the school districts here in the US and Tamil is one of them.

Annual Day

CTA conducts an annual school day at the end of every year. We rent a state of the art auditorium and conduct the events in a professional manner. The entire process of registering for the event, registering kids for the event, uploading of songs and scripts for approval, managing the order and sequence of events are all automated and computerized. Nothing is done manually. We use googlegroups and googledocs to share documents, spreadsheets, create a repository of audio and video files and to automate the approval and registration process. CTA follows strict guidelines to maintain the quality of programs in our annual day. No English words or other language words can be used, no obscene gestures, no vulgar movements and no references to any religion or caste can be used. Only appropriate costumes can be worn by the children.

Annual day is celebrated to add fun to the learning process and also express your cultural awareness in the form of a song, dance or drama. Along with annual day, CTA celebrates Halloween, Deepavali, Thanksgiving, Christmas, Pongal, Tamil new year and conducts a graduation ceremony at the end of the year. These are all fun related activities that add another dimension to the learning process.

Tamil Virtual University

CTA runs TVU classes right after regular school hours are over. Beginner TVU, intermediate TVU, advanced TVU level classes are offered and many of our regular CTA students take TVU classes and get certified by TVU. We use the same system to register our TVU students, maintain their details and track their performance. Many of our students have been certified by TVU at all levels.

Conclusion

CTA's mission and vision of imparting a high quality and streamlined education to Tamils living in the US is made possible by technology. Tamil education to kids born and raised in a foreign land has its own challenges. We try to solve those problems by putting a process in place that is consistent among various branches spread across the entire country. That process implementation and execution is made possible by technology. As we grow in number, we are also strengthening our technology infrastructure to accommodate the volume transactions and queries. The Internet and web based systems are the future for CTA to effectively and smoothly run such a big organization and operation.

தமிழ்க் கற்றல் கற்பித்தலில் மின் கற்றல் முறை

பேரா. ம.செ. இரபிசிங்

1.0 அறிமுகம்

மரநிழல், திண்ணை, வகுப்பறை என்ற எல்லைக்குள், குறிப்பிட்ட கால வரையறைக்குள் வழங்கப்பட்டு வந்தக் கற்றல்முறை, தகவல் தொடர்புத் தொழில்நுட்ப வளர்ச்சியின் பயனால், பரந்து, விரிந்து, எல்லைகளையும், வரையறைகளையும் கடந்து இணையவழிக் கல்வியாக - கற்றலாக இன்று வளர்ந்துள்ளது. உலகளாவிய நிலையில், தமிழ் மொழிக் கற்றலுக்கும் கற்பித்தலுக்கும் உகந்த ஓர் அறிவியல் வழி இணையம். இதன் முதுகெலும்பாய்த் திகழ்வது மின்கற்றல். இக் கற்றல் முறையில் தயாரிக்கப்பட்டத் தமிழ் இணையப் பல்கலைக்கழகப் பாடங்கள் மற்றும் அமெரிக்காவிலுள்ள பென்சில்வேனியாப் பல்கலைக்கழகப் பாடங்கள் பற்றிய ஒரு மதிப்பீடாக இக்கட்டுரை அமைகிறது.

1.1 மரபுவழிக் கற்றலும் இணையவழிக் கற்றலும்

ஆசிரியரையே மையமாகக் கொண்டு, ஆசிரியரை நம்பி, ஆசிரியரே எல்லாம் வல்ல ஆசானாகச் செயல்பட்டு, தனி ஆளுமை செய்து கற்பித்த மரபுவழிக் கற்றல் முறை, கணினியின் பயன்பாட்டால், - குறிப்பாகத் தகவல் தொழில்நுட்ப வளர்ச்சியால் பெற்றப் பல்லுடகத் தொழில்நுட்ப நெறிமுறைகளால் (Multimedia Technologies) இணையவழியில், மாணவன் தானே கற்கும் புதிய கற்றல் முறையாக - மின் கற்றலாகவும் இன்று வளர்ச்சி பெற்றுள்ளது.

2.0 மின்கற்றல்

மின்னியலின் (Electronics) பயன்பாடுகளைப் பயன்படுத்திக் கணினியின் துணையுடன் கற்பிக்கும் 'CSCL (Computer Supported Coloborating Learning) என்று கூறப்படும் கற்றல் முறை 'மின்-கற்றல்' என்று அழைக்கப்படுகிறது. 'E-learning' என்று குறிப்பிடுவதில் 'E' என்பது 'Electronics' என்பதைச் சுட்டினாலும், இக்கற்றலின் தன்மைகளின் அடிப்படையில், Evolving or Everywhere or Enhanced or Extended என்பவையும் பொருந்தும் என்றும் சுட்டுவர். மாணவன் சமயமாகச் சிந்தித்தல், தன்னைத்தானே மதிப்பிடல், தன்னைத்தானே வளப்படுத்தல், குறிகோளுடன் செயல்படல், தொடர்ந்து விருப்பத்துடன் கற்கும் ஆர்வம் பெறல் போன்ற நல் வாய்ப்புகளை நல்கும் இக்கல்விமுறை, கற்றலில் ஒரு புரட்சியை ஏற்படுத்தியுள்ளது என்னும் மிகையாகா.

2.1. மின்கற்றல் முறை

பொதுவாக, வாகனங்களை ஓட்டுவதற்கான பயிற்சியில், அந்த வாகனங்களைப் பெற்று, அந்த வாகனங்களிலேயே பயிற்சி பெறுவர். அப்பொழுது தான் ஓட்டும் வாகனத்தின் செயல்பாடுகளையும், அதைக் கட்டுப்படுத்தும் திறனையும் அறிய முடியும் என்று நம்புகின்றனர். ஆனால், விமானம் ஓட்டுநர் பயிற்சியில், பயிற்சி பெறுபவர்க்கு விமானம் கொடுக்கப்படுவதில்லை. ஓர் அறையில், ஒரு வரையறை செய்த விமான வடிவமைப்பில் பயிற்சியளிக்கப்படுகிறது. விமானத்தில் இருப்பது போன்ற இயந்திர இணைப்புகள் (Console) பொருந்திய, விமான ஓட்டுநருக்குரிய ஒரு காக்கிட் (Cockpit) இருக்கும். விமானத்தில் சிற்றறையில் (Cabin) உள்ள சன்னல்களுக்குப் பதிலாக, இரண்டு திரைகள் பொருத்தப்பட்டிருக்கும். கணினியின் உதவியால், விமானத்தை ஓட்டுவதற்கான அறிவுறுத்தலும், வழிகாட்டல்களும் (Guidance) செயல்முறை பற்றிய விளக்கங்களும் பயிற்சியாளர்கள் பெறுவர். இவற்றின் துணையுடன் பயிற்சியாளர், ஒரு விமானத்தைத் தானே ஓட்டிப் பயிற்சிப் பெறுவதுபோல, உயரம் (Attitude), கட்டுப்பாடு (Control), செய்நிரல் (Programme) ஆகியவற்றைக் கணிக்கும் அறிவும், திறனும், தீர்மானமும் (Decision) தானே

பெற்றுக் கொள்வர். ஓர் அறையில் பொருத்தப்பட்டுள்ள, இந்த அமைப்பில் உள்ள விமானத்தில் அமர்ந்து, விமானத்தைத் தானே இயக்குவது போன்ற அனுபவத்தைப் பெறுகின்றனர். இச் செயற்பாடுகள் முழுவதையும் கணினியின் வாயிலாகவே பெறுகின்றனர். விமானத்தில் அமர்ந்து, விமானத்தை ஓட்டாமலே, விமானத்தை ஓட்டுவதற்குரிய பயிற்சியைத் திறனுடனும், புரிதலுடனும், நம்பிக்கையுடனும் குறிக்கோளுடனும், தன்னைத்தானே மதிப்பிடலுடனும், பெற்றுக் கொள்வதைப்போல, வகுப்பறைக்குச் செல்லாமலே, ஆசிரியரின் வழிகாட்டல் இல்லாமலே, வகுப்பறையில் பெற இயலாத பல நவீன தொழில்நுட்ப வாயிலானப் பல்லாடகங்களின் துணையுடன் கற்கும் கற்றலை மின்கற்றல் வழங்குகிறது.

இந்தியாவில் காண்பூரிலுள்ள இந்திய தொழில்நுட்பக் கல்வி நிறுவனம் (IIT) மின்-கற்றலுக்கான ஒரு தளத்தை அமைத்து, மின்கற்றலை ஊக்குவித்துப் பிரபலப்படுத்தி வருகிறது.

3.0 இணையவழித் தமிழ்ப் பாடங்களும் மின்கற்றலும்

இணையவழியில் தமிழை முதன்மை மொழியாகக் கற்பிக்கும் தமிழ் இணையப் பல்கலைக்கழகப் பாடங்களின் உருவாக்கமும் தமிழை இரண்டாம் மொழியாகக் கற்பிக்கும் பென்சில்வேனியாப் பல்கலைக்கழகப் பாடங்களின் உருவாக்கமும் மின்கற்றலின் அடிப்படையிலானவை. இணைய வழியிலான கற்றலில், மாணவனை மையமாகக் கொண்டு, மாணவனின் கற்றல் சூழல், வாய்ப்பு, தேவை கற்கும் தன்மை, கற்றல் அனுபவம் முதலியவற்றை மனத்திற் கொண்டு மாணவன் விருப்பத்துடனும், புரிதலுடனும் தொடர்ந்து மீண்டும் மீண்டும் கற்று மகிழ்வதையும் வகையிலான மின்கற்றல் முறையில் தமிழ்ப் பல்கலைக்கழகப் பாடங்கள் தயாரிக்கப்பட்டுள்ளன. மழலைக் கல்வி முதல், பட்டயம், மேற்பட்டயம், இளங்கலைப் பட்டம் வரையிலும் வழங்கப்படும் தமிழ்ப் பாடங்கள், படித்தல், பார்த்தல், கேட்டல் ஆகிய தன்மைகள் அமைந்தவாறு, உருவாக்கப்பட்டுள்ளன. எனவே, படங்கள் (Visual pictures), அசைவுப் படங்கள் (Animations), படவரைவு (Graphics), ஒலி (Audio), ஒலி, ஒளிக்காட்சிகள் (Video) ஆகியன பாடங்களில் இடம் பெற்றுள்ளன.

3.1 தமிழ் இணையப் பல்கலைக்கழகப் பாடங்கள்

தமிழ் இணையப் பல்கலைக்கழகப் பாடங்கள் மின் கற்றலுக்குச் சிறந்த எடுத்துக் காட்டாய்த் திகழ்கின்றன. நாட்டியக் கலையில், 'அலாரிப்பு' என்பது ஒரு நடன நிகழ்ச்சி. இதைப் பற்றிய பாடம் பல்கலைக் கழக இணையத் தளத்தில் இடம் பெற்றுள்ளது. பாடத்தில், 'அலாரிப்பு' என்ற நடன நிகழ்ச்சியைப் பற்றிய விளக்கம் தரப்படுகிறது. பின்னர், அலாரிப்பிலுள்ள பாடல், இசையுடன் ஒலிக்கப்படுகிறது. அடுத்து, 'அலாரிப்பு' நடன நிகழ்ச்சி, ஒலி ஒளிக் காட்சியாக இடம் பெற்றுள்ளது. மாணவன் விரும்பினால், நடன நிகழ்ச்சியை மீண்டும் மீண்டும் பார்த்து, நடன பயிற்சியையும் தானே பெறும் வாய்ப்பையும் இப்பாடம் நல்குகிறது.

மேலும் பாடத்தின் இறுதியிலுள்ள காட்சியகத்தில் பாடம் தொடர்பாகவும், நாட்டியம் பற்றியும் அறிந்து கொள்ளும் வகையில் பல்வேறு வகையான நாட்டிய நிகழ்ச்சிகள் ஒலி ஒளிக் காட்சிகளுடன் இடம் பெற்றுள்ளன.

பாடத்தின் இடையில், மாணவன் தன்னை- தன் கற்றலை மதிப்பீடு செய்து கொள்ளும் வகையில் தன்மதிப்பீட்டு வினாக்களும் இடம் பெற்றுள்ளன.

சிலப்பதிகாரம் பற்றிய பாடத்தில் சிலப்பதிகாரத்தின் கதை, ஒலியுடன், சில விநாடிகளில் அசைவுப்படங்களாகக் (Animation) காட்டப்படுகின்றன. கண்ணுக்கினிய காட்சிகளாக வழங்கப்படுவதால், தொய்வில்லாமல், ஆர்வத்தோடும், மகிழ்வோடும் விரும்பிக் கற்கும் சூழலை உருவாக்குவதோடு, கால விரையத்தையும் தவிர்க்கிறது. பல பக்கங்களில் கூறவேண்டிய நெடியகதை அசைவுப் படங்களாகக் காட்சிப் படுத்தப்படுவதால் மனத்தில் பதிந்து, மனக்கண்ணின் முன்னால் வந்து கொண்டிருக்கும் - மனத்திலிருந்து அகலாது.

கணினியின் செயற்பாட்டையும், தொழில்நுட்ப வளர்ச்சிகளையும் பயன்படுத்தி, எத்தனை முறை வேண்டுமானாலும், எப்பொழுது வேண்டுமானாலும் எங்கிருந்தும் கற்கும் வாய்ப்பும் வசதியுமுள்ள இப்பாட அமைப்புகள், மாணவனைச் சோர்பின்றித் தொய்வில்லாமல், ஒருகுறிகோளுடன் விரும்பிக் கற்கும் வாய்ப்புகளை நல்குகின்றன.

3.2 பென்சில்வேனியாப் பல்கலைக்கழகப் பாடங்கள்

இணையவழியாகத் தமிழை இரண்டாம் மொழியாகக் கற்பிக்கும் பென்சில் வேனியாப் பல்கலைக் கழகப் பாடங்களும் மின்-கற்றலின் அடிப்படையிலானவை. இணையப் பல்கலைக் கழகப் பாடஅமைப்பில் படங்கள், படக்காட்சிகள், ஒலி, ஒளிக்காட்சிகள் முதலியன, பாட அமைப்பில் இடம் பெற்றுள்ளன.

பேச்சுத்திறன், எழுதும் திறன், படிக்கும் திறன், மற்றும் புரிந்து கொள்ளும் திறன் ஆகிய நான்கு திறன்கள் இரண்டாம் மொழியாக ஒருமொழியைக் கற்கும் மாணவர்களுக்கு அவசியமாகிறது (வாசு. அரங்கநாதன். Tamil Internet சிங்கப்பூர் 2003). இவற்றின் அடிப்படையில் பென்சில் வேனியாப் பல்கலைக் கழகப் பாடங்கள் உருவாக்கப்பட்டுள்ளன.

இப்பல்கலைக்கழகப் பாடங்களைக் கற்பதற்கான வசதிகளும் வாய்ப்புகளும் மாணவன் தானாகக் கற்பதற்காக உருவாக்கப்படவில்லை, என்றாலும் அவற்றைத் தானாகக் கற்கும் மாணவனும் பயன்படுத்திக் கொள்ளும் வகையில் தயாரிக்கப்பட்டுள்ளன என்று பாடத்தைப் பற்றிய அறிமுகவுரையிலேயே கூறப்பட்டுள்ளது.

சிக்காக்கோ பல்கலைக் கழகத்துடன் இணைந்து உருவாக்கப்பட்ட, இந்த இணையத் தளத்தில், ஆங்கிலத்தின் துணையுடன் மொழியியல் அடிப்படையில், பாடங்கள் உருவாக்கப்பட்டுள்ளன.

ஒலி ஒளி மற்றும் உரை இணைந்த பாடங்கள் கற்போரின் படிக்குத்திறன், புரிந்துகொள்ளும் திறன் ஆகியவற்றை மனத்திற்கொண்டு அமைக்கப்பட்டுள்ளன.

தமிழ் அகர வரிசையை எவ்வாறு எழுதுவது, சொற்களை எவ்வாறு திரும்பத் திரும்பச் சொல்லிப்பார்ப்பது போன்றவற்றிற்கு வெவ்வேறு குறியீடுகளை (பேனாவும் கையும், ஒலிபெருக்கி) அமைத்துத் தானே கற்கும் மாணவன் அவற்றை எளிமையாகப் புரிந்து கொண்டு கற்கும் வசதியுள்ளது.

உரையாடல்கள், குறைந்த அளவு எளிய சொற்களைக் கொண்டு சிலவரிகளிலேயே அமைக்கப்பட்டுள்ளன. இது, தமிழை இரண்டாம் மொழியாகக் கற்கும் மாணவர்களுக்கு சிரமமில்லாமல், எளிதாகப் புரிந்து கொண்டு கற்பதற்கு வாய்ப்பளிக்கிறது.

கற்கும் மாணவர்கள், தமிழை எளிமையாகக் கற்றுக் கொள்ளும் வகையில் தமிழ்ச் சொல் அட்டவணை, உரோமன் எழுத்துகளிலும், ஆங்கில மொழிபெயர்ப்புகளிலும் வழங்கப் பட்டுள்ளன.

கற்றலை மீண்டும் பரிசீலனை செய்துபார்க்கும் வகையில் பாடஅமைப்புகள் உள்ளன.

4.0 இணையத்தளப் பாடங்களின் அமைப்பு

கல்வியாளர்கள், கல்வி கற்பதற்கு ஒரு சூழல் தேவை என்பர். அதனை வகுப்பறைச் சூழல் (Classroom climate) என்பர். வகுப்பறைச் சூழல் என்றால், ஆசிரியர் கூறுவதைக் கேட்கும் மனநிலை, கேட்கும் பாடத்தில் விருப்பம் அல்லது ஈடுபாடு. மேலும் அறிந்து கொள்ளவிழையும் ஆர்வம் ஆகியவை மாணவர்க்கு ஏற்பட வேண்டும் என்பர். மாணவரின் உளவியலை அறிந்து, மாணவருடன் நேருக்கு நேர் தொடர்பு கொண்டு நெறிப்படுத்தும் தன்மையைத் தவிர, ஏனைய மேற்குறிப்பிட்டக் கூறுகளை மனத்திற்கொண்டே தமிழ் இணையப் பல்கலைக் கழகப் பாடங்கள் தயாரிக்கப்பட்டுள்ளன. மாணவர்களுடன் நேரடியாகத் தொடர்பு கொண்டு, பாடம் தொடர்பான ஐயங்களைத் தெளிவுப் படுத்துவதோடு அவர்களது மனநிலையையும் அறியும் வகையில், பாட ஆசிரியர்களோடு அல்லது தொடர்புடையவர்களோடோ மின்-அஞ்சல் அல்லது

தொலைபேசி மூலம் தொடர்பு கொள்ளும் வசதிகள் தமிழ் இணையப் பல்கலைக் கழக முதல் இயக்குநர். பேராசிரியர். மு. பொன்னவைக்கோ அவர்களால் வழங்கப்பட்டிருந்தது. அவ்வாய்ப்பு மீண்டும் புதிப்பிக்கப்பட்டால், மேற்குறிப்பிட்ட வகுப்பறைச் சூழல் வளம் பெறும். பென்சில்வேனியப் பல்கலைக்கழகத்தில் இவ்வசதியுள்ளது.

இன்றைய இணையவழிக் கல்வி வளர்ச்சியில், இணைய வழிக் கலந்துரையாடல் (E-chat), இணைய வாயிலான ஒலி, ஒளிச் சந்திப்பு (Video Conference), இணைய வகுப்பறை (Virtual Classroom) போன்ற வசதிகளும் வாய்ப்புகளும் உள்ளன. அவற்றையும் இணைய வழியில் கற்கும் இணைய பல்கலைக் கழக மாணவர்களுக்கு அறிமுகப்படுத்த வேண்டும். இன்றைய இணையவழிக் கல்வியில் அவை தவிர்க்க இயலாதவை. இதனால் இணைய வழிக் கற்றலை மேம்படுத்தலாம்.

பாடத்தைக் கற்கும் பொழுது, ஏதாவது ஒரு சொல்லுக்குப் பொருள் தெரியாவிட்டால், தேடுதல் வசதியின் மூலம் அகராதியிலிருந்து உடனே பொருளைப் பெறும் வாய்ப்பையும் ஏற்படுத்தலாம். பாடம் தொடர்பாகவும், கற்கும் மாணவர் தம்மை வளப்படுத்தவும் பாடத் தொடர்புடைய வேறு வலைத் தளங்களின் இணைப்புகளைப் பாட இறுதியில் தரலாம்.

கூகுல் தேடுதல் பொறியின் வசதியைப்போல், பாடம் தொடர்பாக, ஒன்றிலிருந்து இன்னொன்று என்று தொடர்ந்து மாணவன் தேடிப் பெறும் வாய்ப்பும் மாணவனுக்குப் பெருந்துணையாக அமையும்.

இக்கற்றல் முறையில் பாடங்களுக்கும் பாட இணைப்புகளுக்கும், நினைத்த உடனேயே எளிமையாகச் செல்ல முடியாதத் தொழில்நுட்பக் குறைபாடுகளைத் தவிர்த்தால், கற்கும் மாணவன் தடையின்றி விருப்பத்துடன் கற்க இயலும்.

5.0 தொகுப்புரை

கணினியின் செயற்பாடுகளால் பெற்ற வாய்ப்புகளையும், வசதிகளையும் பயன்படுத்திப், பரந்துபட்டு வளர்ந்து வரும் கல்வித் துறையில், இணையவழிக் கல்வி என்பதுவும், அதன் அடிப்படையான மின்கற்றல் என்பதுவும், இன்றையச் சூழலில் ஓர் அரியவாய்ப்பு, தவிர்க்க இயலாத ஒன்று, வளர்ச்சிக்குத் தேவையான ஒன்று, தகவல் தொடர்புத் தொழில்நுட்பப் பங்களிப்புகளில் ஒன்று. தமிழ்மொழியை உலகளாவிய நிலையில் கற்பிப்பதற்கும், பரப்புவதற்கும் பெருமளவில்துணை செய்யும் ஒன்று.

மாணவர்கள், ஆசிரியர்களுடன் நேரடியாகத் தொடர்பு கொண்டு கற்கும் வகுப்பறைக் கற்றலுக்கெனச் சிறப்புக் கூறுகள் பல உண்டு. மின்கற்றல் என்பது அறிவியல் வளர்ச்சியின் வெளிப்பாடு. ஆற்றல் வாய்ந்த ஒரு கற்றல் முறை. இன்றையச் சூழலிற்கேற்ற ஒரு புது வழி. காலத்தையும், பொருளையும் மிச்சப்படுத்தும் ஒரு நவீன கற்றல்முறை. உலகு தழுவிய வாழும் தமிழர்களும், தமிழ் ஆர்வலர்களும் தமிழ் மொழியைக் கற்பதற்குரிய ஒரு சிறந்த வாயில்.

A Study on the Role of Tamilvirtual University in Tamilteaching and Learning at Elementary Level

Dr.Nirmala Devi.S

Associate Professor & Head, Dept.of Education,
Govt.Institute of Advanced Study in Education,
Chennai.15 nirmaladevielango@gmail.com

Dr.Rajeswari.T

Research Fellow,
ECOLE FRANCAISE,
Pondicherry

Introduction

Education is a process of human enlightenment and empowerment for the achievement of a better and higher quality of life. Education develops the total personality of an individual. It also contribute to the growth and development of the society. The field of education is expanding each year as advancement is made in technology and brain based research.

Technology and Education

New developments and new technologies came into existence and any growing nation cannot remain away from the needs to change in time so that the society is not labeled as under developed and backward society. The process of education cannot ignore the social and psychological impacts of the technology that structures information

Importance of Teachers

Teacher is a significant agent in causing learning and intellectual development of the learner. It is necessary that teachers keep themselves abreast of new developments. **Educators** continually strive to maximize the effectiveness and efficiency of teachers by identifying and comparing alternative method of teaching and learning.

Changing Role of the Teacher

During the era of Information Technology the teacher can be called as a mentor, monitor and motivator ie to develop better learning styles and information seeking behaviour among children. Hence during the 21st century teacher's new role is that of a facilitator and designer of learning experiences and to maintain innovations in the classroom.

Need for the study

The course of Tamil Virtual University is available free of cost.

Any new technology comes not merely with hardware and software but with a learning and teaching style and grammar of its own, and that management practices need to be adapted in order to use the technologies effectively.

With poor access and high digital divide teachers tend to resist adoption of e-learning.

To develop teacher competencies in use of web technologies, to plan for capacity building and Training, the teacher's awareness about the functioning of TamilVirtual University, its various teaching modules, their attitude towards the usage of web based technology need to be assessed. This will help the planners and administrators of education to plan and implement in service program to teachers and to make certain changes in curriculum, to make recommendations to the govt the need for improving the infrastructure facilities and technical support in schools.

Need for Technology

Web based learning is an essential tool for achieving sustainability and will help in enabling better and increased access to information to enrich the teaching learning process.

Tamil Virtual University

The **Tamil Virtual University** was established in 2000 by the Government of Tamilnadu, aims at providing Internet based resources using multimedia and opportunities for the Tamil communities living in different parts of globe as well as others interested in learning Tamil and acquiring knowledge of the history, art, literature and culture of the Tamil. This is the first virtual university in India for language(Tamil) teaching. Similarly the Indira Gandhi National Open University(**IGNOU**) started virtual campus initiatives (**1999**), **Netvarsity (1996)** first online learning facility by NIIT, **Yashwantrao Chavan Maharashtra Open University** are some of the other virtual universities offering courses through on line mode. **Newport Asia Pacific University (NAPU)** a new virtual university that offers programs to teach Japanese as a second language. Nunnan, David(2002)

The **TVU** web based course is designed and given in four levels viz

- 1.Pre Primary Level
- 2.Basic Level
- 3.Certificate Intermediate Level
4. Certificate Advanced Level

The course material is also available in CD. The certificate course is recognized by the govt. of Tamilnadu. The content given in the CD includes an introduction, lessons related to listening, reading skills, introduction of grammar components, and related exercises, follow up activities, self evaluation. .

Objectives of the study

The main objective of the present study is to find out

How effectively this web resource is utilized for Tamil teaching-learning processes at the elementary level by the Tamil teachers

Which part and which level of the content given in the web is being widely used by the teachers

Their attitude towards the usage of this web resource

Their willingness to learn the fundamentals of integrating technology in classroom, type of response and interest shown by the students in using web resources

Whether students are willing to use this web resources for improving their language skills,

The encouragement and motivation given to the teachers by the Govt. and Management, the infrastructural facilities available in the school premises

Method and Procedure

The present study employed Descriptive Survey research method.

Sample

190 Primary School Tamil Teachers (in and around Chennai and Puducherry) constituted the sample for this study.

Tool

The researchers constructed a questionnaire with 22 statements on a 3 point scale under four dimensions (Yes, No, At times) for this study. (Sample tool is given in Table1)

Analysis of Data

The collected data is quantified and the findings are given as percentage for the responses related to the option -Yes .

Findings of the Study

78% of the sample were aware of the establishment and functioning of TVU, 69% were familiar with the web site address. 60% have expressed that they made attempts to know about the content developed by TVU. 66% agreed that the visuals given for writing practice is very useful. 89% confirmed that their students are interested to learn Tamil through the web using Multimedia technology.86% stated that learning Tamil through this virtual mode is very useful for the students. 77% were of the opinion that this virtual mode has made teaching and learning of Tamil easy.

75% agreed that lessons given at the intermediate level do develop language skills of the students. Only 50% of the sample have stated that they have the relevant facilities in schools to utilize this web resources, 50% thought of using web based technology.

53%, 66%, 77% have agreed for the statement 16, 17, &18 respectively.

77% confirmed that poetry lessons were very interesting. 90% accepted the need for change in the teaching methodology.61% were of the opinion that there are differences between traditional methods and web based multimedia technology. 84% are willing to undergo training in web based teaching methodology. 90% wanted training in all aspects to use this web based technology for teaching learning purposes.

Educational Implications of the Study

1. Intensive basic training program, periodical refresher courses must be planned and implemented.
2. The web based learning material CD can be given to all schools.
3. Technology based language teaching to be introduced in the primary level school curriculum .

Conclusion

Since 90% of sample studied want change in the teaching methodology and 89% students are interested in learning through this method if proper facilities and support services along with effective training are provided tamilthai will get a new face in the 21st century.

References

1. **Anantha Sayanam.et al.** Multimedia as an alternative strategy in Teaching-Learning process in Higher Education. The educational Review. Dec.1998.
2. **Arulsamy.S(2010)** Educational Innovations and Management. Neelkamal publications. Hyderabad.
3. **ERIC WEB Sources.**
4. New Trends in Pedagogical Techniques. M.Phil Education Tamil Nadu Open University, School of Education.
5. **Panday.K.P.** E-learning: CONCEPT, POTENTIAL AND FUTURE. Indian Journal of Teacher Education. ANWESHIKA. 5(1), June 2008. NCTE, New Delhi.

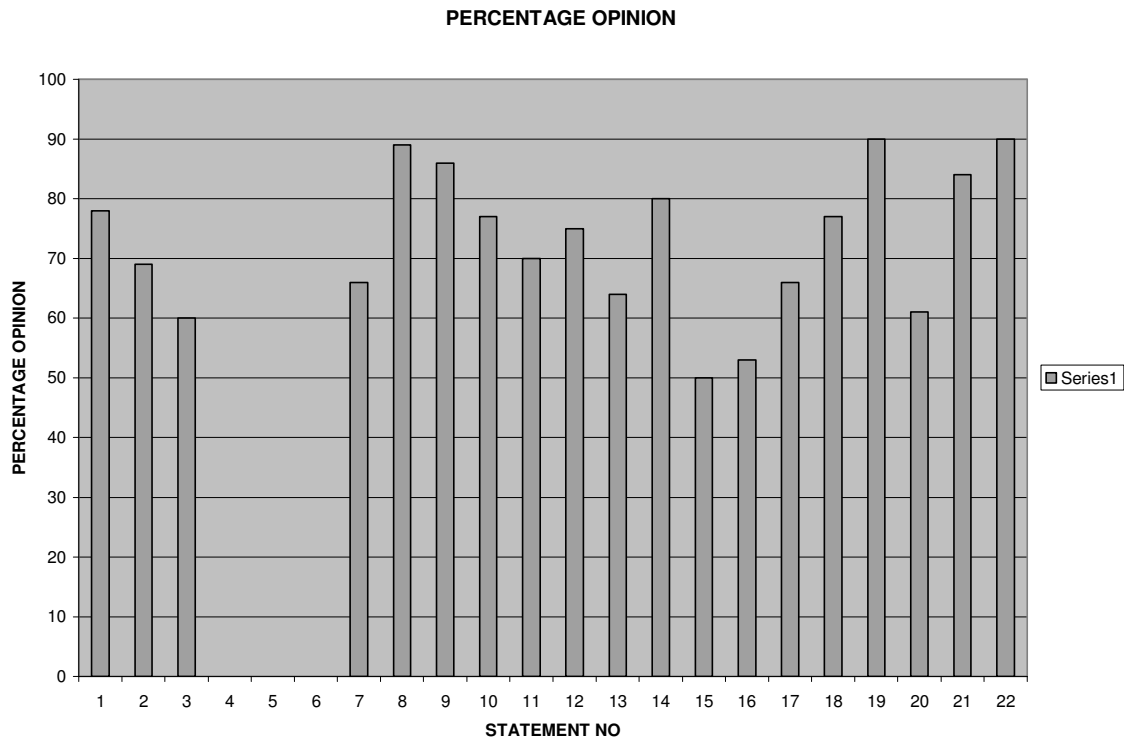


TABLE 1

SNO	DESCRIPTION	Yes/ No/ At times
1	Are you aware that the Tamil Virtual University established by the Govt of Tamilnadu has been functioning?	
2	Have you come across the website of Tamil Virtual University : www.tvu.org	
3	Have you made any attempt to know that the TVU has prepared lessons for school students?	
4	As of now teaching is done in 4 stages in TVU. Mark the stage which you were aware of 1.Primary Education Level 2.Basic Level 3.Certificate Intermediate Level 4. Certificate Advanced Level	
5	Which among the following lessons of Primary Education you have used? words, songs, events, conversation, stories, numbers, letters	
6	Indicate which among the following lessons given for basic level in TVU that you are aware of 1.lesson on Tamil letters 2.Group of Tamil letters 3. Writing Practice 4.Kirantha Eluthukkal 5. Song on introduction of yureluthukkal	
7	Is the visuals on writing practice given in the web easy to follow?	
8	Whether students are interested to learn lessons through this audio visual mode?	
9	Is teaching through AV MODE is beneficial to the students?	
10	Is teaching through AV mode make the process of teaching easier?	
11	Whether students are interested to learn through multimedia?	
12	Is the lessons meant for intermediate level enhances the language skills of the students?	
13	Do the illustrations given for vatrumai urubu make the students to learn grammar with interest and willingness?	
14	Are the students willing to learn through the web?	
15	i) Do you have any idea of teaching through web to improve the language skills of your students ? ii) Do you have the facility to utilize web for improving the language skills of your students in your school ?	
16	Do you consider the lessons of higher level in the web are equivalent to VI std of formal school level?	
17	Is the method of practice given in the higher level is easy to learn?	
18	Is the poetry lessons given in the web motivate learning ?	
19	Do you agree that the teaching learning methods to be changed according to the changing time?	
20	Is there any difference between the traditional and multimedia methods of teaching in enhancing the language skills of students ?	
21	Are you interested in undergoing training programs to use web based teaching methods in your class?	
22	What sort of training you require?	

மலேசியத் தமிழரல்லாதார்க்குக் கணினி வழி

தமிழ்மொழி: ஒரு பார்வை

Paramasivam Muthusamy
University Putra Malaysia
param@fbmk.upm.edu.m

கட்டுரைச் சாரம்

மலேசிய புத்ரா பல்கலைக்கழக மொழியியல் துறையில் தமிழ்மொழி (தொடக்க நிலை) கற்பிக்கப்பட்டு வருகிறது. தேர்வுப் பாடங்களின் வரிசையில் தமிழ்மொழியும் ஒரு விருப்பப் பாடமாக இருந்து வருகிறது. பிறமொழி மாணவர்கள் குறிப்பாக மலாய் சீன மாணவர்களில் சிலர் தமிழ்மொழியைப் படித்து வருகின்றனர். சுப.திண்ணப்பன் (1992), தமிழ் புழங்கும் புறச் சூழல் அற்ற இடத்தில் பிறமொழியாளர்களுக்குத் தமிழ்மொழி கற்பித்தல், அயல்மொழி கற்பிக்கும் முறை என்பது போல புத்ரா பல்கலைக்கழகத்தில் அயல்மொழி முறையிலேயே தமிழ்மொழி கற்பிக்கப்பட்டு வருகிறது. தகவல் தொடர்புத் தொழில் நுட்ப வளர்ச்சிக்கு ஏற்றவாறு இன்று தமிழ்மொழியையும் கற்பிக்க வேண்டும். கற்றல் கற்பித்தல் நடவடிக்கைகளைக் கணினி வழி கையாளுவதால், தமிழரல்லாதவர்களிடம் தமிழ் கற்கும் ஆர்வம் மேலோங்குகிறது. ஆசிரியர் துணையின்றித் தானே சுயமாக கற்க்கின்றனர். ஆகவே புத்ரா பல்கலைக்கழக மலாய்சீன மாணவர்களுக்குக் கணினி மென்பொருள் , இணையம் , வலைப்பக்கம் , இணையப் பல்கலைக்கழகப் பாடங்கள் போன்றவைகளின் பயன்பாடு எங்ஙனம் உதவுகிறது என்பதைக் கட்டுரை ஆராய்வதோடு சில பரிந்துரைகளையும் முன் வைக்கின்றன.

மூன்றுரை

மலேசிய புத்ரா பல்கலைக்கழக மொழியியல் துறையில் தமிழ்மொழி (தொடக்க நிலை) கற்பிக்கப்பட்டு வருகிறது. பிறமொழி மாணவர்கள் குறிப்பாக மலாய், சீன மாணவர்கள் தமிழ்மொழியை விருப்பப் பாடமாகத் தேர்ந்து படித்து வருகின்றனர். இவ்வகுப்பு ஒவ்வொரு பருவத்திற்கும், (semester) ஒரு நிலை (level) என மூன்று பருவங்கள் வரை தொடர்கின்றது.

தமிழ் கற்பித்தலை முதல் மொழி நிலை, இரண்டாம் மொழி நிலை, அயல் மொழி நிலை என்று மூன்று வகையாகப் பகுத்து காண்பர். தமிழைத் தாய் மொழியாய்க் கொண்டவர்களுக்குத் தமிழ் கற்பித்தல் முதல் மொழி நிலை என கருதப்படும். தமிழைத் தாய்மொழியாகக் கொள்ளாத பிற மொழியைச் சார்ந்தவர்களுக்குத் தமிழ் கற்பித்தல் இரண்டாம் மொழிநிலை எனப்படும் தமிழ் புழங்கும் புறச் சூழல் அற்ற இடத்தில் பிறமொழியாளர்க்குத் தமிழ் கற்பித்தல் அயர்மொழி நிலையில் வழங்கப்படும் (சுப.திண்ணப்பன், 1992)

புத்ரா பல்கலைக்கழகத்தில் அயல்மொழி என்னும் நிலையிலேயே தமிழ் கற்பிக்கப்படுகிறது. அயல்மொழி நிலையில் கணினி வழி தமிழ்மொழி எவ்வாறு கற்பிக்கப்படுகிறது என்பதை இக்கட்டுரை காட்டுகிறது.

வாசிப்பு, கேட்டல், எழுத்து, பேசுதல் போன்ற மொழிக் கூறுகள் கணினி துணைக் கொண்டு கற்பிக்கப்படுகிறது.

வாசிப்பில் கணினியும் காட்சியும்.

மொழித்திறனில் வாசிப்புத் திறனை வளர்ப்பதில் கணினி பெரிதும் பயன்படுகிறது .வாசிப்பைத் தொடங்கும் முன், வாசிப்புப் பனுவலின் கருப்பொருளோடு தொடர்புடைய படங்களையோ, படக் காட்சிகளையோ கணினியில் சேமித்து எடுத்துச் சென்று வகுப்புத் திரைகளில் காட்டலாம். இதன் மூலம் மாணவர்கள், வாசிப்பின் பால் முழு கவனத்தைச் செலுத்துவர்.

இவ்வாறு படக்காட்சிகளையோ படங்களையோ காட்டும் போது மாணவர்களுடன் அதன் தொடர்பாகப் பேச வேண்டும். பேசும் பொழுது தமிழ் சொற்களுக்கு ஆங்கிலத்தில் விளக்கம் தர வேண்டும்.

எ.கா. அணில், அம்மா, ஆடு, இலை என்று திரையில் படம் வரும்பொழுது உரக்கச் சொல்ல வேண்டும்.

அணில்

படத்தைப் பார்த்தவுடன் பிறமொழி மாணவர்கள் அதைத் தங்கள் மொழியில் புரிந்துக் கொள்வர். ஆங்கிலத்திலும் அதன் பெயரை கூற வேண்டும்.

படத்தைக் காட்டியதன் தொடர் நடவடிக்கையாக காட்டப்பட்ட படத்தின் எழுத்துகளைத் திரையில் தனித் தனியாக விழச் செய்ய வேண்டும்.

அணில்

இதன் மூலம் மாணவர் எழுத்தையும் அதன் ஒலியையும் ஏக காலத்தில் அறியும் வாய்ப்பு ஏற்படுகின்றது. இம்மதிரியான படக்காட்சி மற்றும் படங்களிணைந்த வட்டுகளை நிறைய தயார் செய்வது என்பது எளிதல்ல. TISC என்று சொல்லக்கூடிய தமிழ் இணைய ஒருங்கிணைப்புக் குழுவினருடன் இணைந்து கணினி வட்டுகளைச் சொந்தமாகத் தயாரிக்கலாம்.

உச்சரிப்பு

தமிழ் எழுத்துக்களின் ஒலியமைப்பு தமிழல்லாத மாணவர்களுக்கு உச்சரிப்பு நிலையில் சிரமத்தைத் தருகின்றன. குறில் நெடில் வேறுபாடு, வல்லின, மெல்லின. இடையின எழுத்துக்களின் உச்சரிப்பில் அவர்கள் சிரமத்தை எதிர்நோக்குகின்றனர்.

குறிப்பாக ல,ள,ழ,கர எழுத்துக்கள், ற்,ற் ஆகிய எழுத்துக்களின் உச்சரிப்பு த், ட் உச்சரிப்பில் ஒலி வேறுபாடு தமிழரல்லாத மாணவர்களின் நாவிற்கு கட்டுபட மறுக்கிறது. இதற்குப் பொருத்தமான ஆங்கிலச் சொற்களின் வழியாக ஏற்ற உற்சரிப்புகளைக் கட்டாயம் கூற வேண்டும்.

மூக்கொலிகளான ங்,ஞ்,ண்,ந்,ன ஆகிய மெல்லின மெய் ஒலிகள் சிரமத்தைத் தருகின்றன. இதற்கு ng,ny,n ஆகிய மலாய் மொழி உச்சரிப்பைத் துணையாகக் கொள்ளலாம். ஒலிநாடா வழி அல்லது கணினியில் இவ்வெழுத்துக்களைக் காட்டி ஏற்ற உச்சரிப்பு ஒலி வேறுபாட்டிணைத் தெளிவாக கூற வேண்டும்.

சில ஆங்கில மொழி வாக்கியங்கள் அப்படியே நேரடியாக தமிழில் உச்சரிக்கும் பொது பொருள் வேறுப்பாட்டிணை ஏற்படுத்தும்.

எ.கா: He eat with a spoon.

அவன் கரண்டியோடு சாப்பிட்டான்.

இவ்வாறு நேரடி மொழி பெயர்ப்பு தவறான பொருளை ஏற்படுத்தி விடும். மேலும் மிக நீண்ட வாக்கியங்களை கற்பிக்காமல் சுருக்கமான வாக்கியங்களையே கற்பிக்க வேண்டும்.

தொடரியலிலும் சில சிரமங்கள் உள்ளன. எழுவாய், பயனிலை இயைபு (தினை,பால்,எண்,இடம்) கற்பதில் சிரமங்கள் ஏற்படுகின்றது. 'He has a book' என்பது 'அவன் ஒரு புத்தகம் வைத்திருக்கிறான்' எனக் கற்கும்

மாணவர்கள் 'He has a son' என்பதை 'அவன் மகன் வைத்திருக்கிறான்' என எழுத வாய்ப்புள்ளது. ஆகவே, உயர்திணை, அஃறிணை வேறுப்பாட்டை தெளிவுபடுத்தும் கடமை ஆசிரியருக்கு உள்ளது.

கணின வழி தமிழ் கற்பிக்கும்போது இவ்வாறான கூறுகளையும் மனதில் கொண்டு செயல்பட வேண்டும்

பாடல்கள்/ இசை நாடா

'மகுடிக்கு மயங்காத பாம்பும் பாடலுக்கு மயங்கா மனமும் இல்லை' எனலாம். தமிழர் அல்லாதவருக்குத் தமிழின் பால் ஓர் ஈர்ப்பை முதலில் ஏற்படுத்த வேண்டுமெனில் அதற்கு பாடல்களே சிறந்த வழிக்காட்டி.

முதலில் எளிமையான பாலர், ஆரம்பப் பள்ளி பாடல்களை அறிமுகப் படுத்தலாம். இன்று சந்தையில் நிறைய பாடல்கள் உள்ளன. சில குறுந்தட்டுகளில் பாடல்கள் 'ரோமன்' செய்யப்பட்டுள்ளன.

எ.கா:

பனிக்கூழ் பனிக்கூழ் பார் பார் பலவிதம்

panikul panikul par par balavitham

இப்படி 'ரோமன்' செய்யப்பட்டுள்ள பாடல்களைக் கேட்டு பாடிப் பழகுவது மிகவும் சுலபம். நமது கடமை முதலில் அவர்களுக்கு தமிழின் பால் ஆர்வத்தைத் தூண்டுவதாக இருக்க வேண்டும்.

ஒருவர் காதால் கேட்பதை விட கண்ணால் பார்த்து, காதால் கேட்டு, வாயால் பாடும் போது அது சுலபத்தில் மனதில் பதியும்.

பிறகு, முரசு நெடுமாறன், அழ வள்ளியப்பா போன்றோரின் பாலர் பாடலை அறிமுகப்படுத்தலாம்.

சுயமாகக் கற்றல்

கணினி கூடத்தில் மாணவர்கள் சுயமாகவே கணினியை இயக்கி கற்கச் செய்வது மேலும் சிறந்த பயனை அளிக்கும். மாணவர்கள் ஒரு விளையாட்டு முறையில் கணினி வழி தமிழ் கற்க வழி செய்யலாம். இதன்வழி மாணவர்கள் தங்களுக்குப் பிடித்தவற்றை மீண்டும் மீண்டும் தட்டிப் பார்த்து கற்க முடியும். மகிழ்ச்சியுடனும் ஆர்வமுடனும் கற்க முற்படுவர்.

பயிற்சிகள்

கணினி மூலமே எளிதான பயிற்சிகளைச் செய்ய பாடத் திட்டங்களைச் செய்ய வேண்டும். உதாரணமாக:

- I. சரியான எழுத்தைப் பொருத்துதல்
—ந்து இ—
- II. சொல்லுக்கு எற்ற படத்துடன் இணைத்தல்
குடை பழம் மலர்

இவ்வாறான, ஏராளமான பல்வகைப்பட்ட பயிற்சிகளைச் சிந்தித்து கணினியின் வழி பயிற்சி பெற ஏற்பாடு செய்யலாம்.

மென்பொருள்கள்

தமிழில் கற்பதுக்குப் பல மென் பொருள்கள் உருவாக்கப்பட்டுள்ளன. இம் மென்பொருள்களைப் பின்வருமாறு வகைப்படுத்தலாம்.

- i. மின்மினி உலகம்; முகாமில் மூன்று நாள்
- ii. வளர்ச்சிப்படி நிலை அமைந்த பாடாத்திட்டம் சார்ந்த மென் பொருள்கள்: செந்தமிழ் 1A

செந்தமிழ் 1B

செந்தமிழ் 4A

iii. கதை அடிப்படையில் மென்பொருள்கள்: தமிழ் சுரபி

ஆரத்தியோடு கதை நேரம்

ஆரத்தி விரும்பும் கதைகள்

iv. இரண்டாம் மொழியாகத் தமிழைக் கற்க ஏதுவான மென்பொருள்கள் : தமிழ் உலகம்

தமிழுக்குத் தயாரிக்கப்பட்டுள்ள விசைப் பலகை மேனேஞர்கள் (Keyboard Managers) பல உள்ளன. அவை

i. கணியன் 2000

ii. சுரபி 2000

iii. தாரகை

iv. முரசு அஞ்சல் 2000

v. தமிழ் 2000

vi. இளங்கோ தமிழ் 2000

vii. ஐ.எஸ்.எம் (ISM)

viii. ஸ்ரீலிபி

ix. விங்க்கி (WINKEY)

பெரும்பாலும் பயன்படுத்தப்படுகின்ற ஆங்கில மென்பொருள்களோடு மேற்சொன்னவற்றுள் கணியன் 2000ஐத் தமிழ் கற்பித்தல், கற்றலுக்குச் சோதித்துப் பார்த்த போது இது ஓரளவு கற்றல் கற்பித்தல் செயல்பாட்டுக்குப் பயன்படுவது புலப்படுகிறது.

ஆசிரியர்கள், மாணவர்களின் பயன்கருதி பிரபலமான ஆங்கில மென்பொருள்கள் தமிழ் கற்பித்தலுக்கு உதவுகின்றன. அவை:

ஆசிரியர்களுக்கு

- MS. WORD
- MS POWER POINT
- MS FRONT PAGE
- MACROMEDIA DIRECTOR
- HOT POTATO

மாணவர்களுக்கு

- STORY BOOK WEAVER DELUXE
- KID WORKS DELUXE
- 3D MOVIE MAKER
- NETSCAPE
- MS WORD
- MS POWER POINT

எதிர்நோக்கும் பிரச்சனைகள்

முதலில் எழுத்து தமிழ் கற்பிப்பதா? அல்லது பேச்சுத் தமிழ் கற்பிப்பதா?

- எழுத்துத் தமிழ் கற்பிக்க ஒரு வரையறை இருப்பதால் இது சுலபமாக இருக்கிறது.

ஆனால், பேச்சுத் தமிழுக்கு அம்மாதிரியான இலக்கணம் கிடையாது. பேச்சுத் தமிழை ஒரு பொது நிலையான பேச்சு மொழியாக (Standard Spoken Tamil) கற்பிப்பதே சுலபமாகும்; சிறப்பாகும்.

ஆக, பேச்சு வழக்கினை கற்பித்த பின்னரே எழுத்து வழக்கிற்குச் செல்ல வேண்டும். ஆங்கில மொழிச் சொற்களைத் தமிழ்ச் சொற்களுக்குப் பதிலாகத் தாராளமாகப் பயன்படுத்த வேண்டும்.

முடிவு

மலாய் சீன மாணவர்களிடம் மெம்மேலும் தமிழ்மொழி கற்கும் ஆர்வத்தை வளர்க்கும் வண்ணம், மலேசியச் சூழல் நிறைந்த கணினி வழி தமிழ் கற்கும் வகையில் - வாசிப்பு , மொழிப் பயிற்சி/விளையாட்டுக்கள், பாடல்கள், கதைகள் மேலும் உருவாக்கப்பட வேண்டும்.

துணைக் குறிப்புகள்

1. திண்ணப்பன், சுப., ஐந்தாம் உலகத் தமிழாசிரியர் மாநாட்டு மலர் (1992)
சிங்கப்பூர்த் தமிழாசிரியர் சங்கம், சிங்கப்பூர்.
2. ஏழாவது உலகத் தமிழாசிரியர் மாநாடு. மாநாட்டு ஆய்வடங்கல் (2006)
கோலாலம்பூர், மலேசியா
3. தமிழ் இணையம்,(2002) மாநாட்டுக் கட்டுரைகள், San Francisco.
4. தமிழ் இணையம்,(2003) மாநாட்டுக் கட்டுரைகள், சென்னை. இந்தியா.

ICT for Tamil education in Tamilnadu

Current Challenges and Opportunitites

Prof S.Balaji,

D.B.Jain College (Autonomous), Chennai-97

Mobile:9840494643

srisaibalaji@rediffmail.com

Introduction

Tamil is a Dravidian language spoken predominantly by Tamil people of the Indian subcontinent. It has official status in the Indian state of Tamil Nadu and in the Indian union territory of Puducherry. Tamil is also an official language of Sri Lanka and Singapore. It is one of the twenty-two scheduled languages of India and the first Indian language to be declared as a classical language by the government of India in 2004. Tamil is also spoken by significant minorities in Malaysia, Mauritius and Réunion as well as emigrant communities around the world.[1].The use of technology for Tamil education for students is very minimal. This is due to lack of confidence, demand for the use of English in the application of technology. Even though the students appreciate the application of technology in learning and teaching of language, these two reasons prevent them from active application of technology .

Driving forces

Considering the fact that English is becoming the dominant home language in most Indian households, there is a need to do more to help children from households who have little exposure to Tamil. Therefore, there is a need to review how we teach Tamil,especially at the primary level, so that our students do not lose interest in the language. In particular, teachers need to concentrate on teaching oral communication skills to the younger generation so that they can communicate in the language more confidently, effectively and in greater depth, and will be motivated to use Tamil within and outside of school. Our children have grown up very comfortable with technology - they use mobiles phones, play computer games and surf the net - we should tap on their IT literacy as well as the excellent infrastructure in schools to teach the languages.The illiterate, the physically challenged and the facility-challenged, all of them need some support or the other to be accepted in society and enjoy the fruits. The traditional methods and practices are invariably driven by us, human beings, and therefore tend to be biased.

The introduction of computer (also known as ICT or IT) has changed the scenario to that of an interactive, collaborative environment where the quest for information knowledge is created actively by students. It is here that ICT can help by providing independence, flexibility and variety to the “less privileged learner”.This is yet another virgin field, where ICT could have a lasting impact, in terms of enhancing the teaching and learning capabilities, respectively, of the agents of change (teachers) and the beneficiaries (the students).

Impact of ICT

Computer-aided education was initially introduced in India as an innovative activity under the District Primary Education Programme (DPEP) in 1994. ICT entered the education sector through DPEP, which spearheaded the design of school information systems. Literature on the role and efficacy of ICT in education is replete with insightful studies. Resnick (2002) opined that computer was akin to finger-painting. In the realm of learning, technology could be employed by students for “making” things, i.e., usage of technology to design and build things of importance. This would consequently increase the acceptance and adoption of technology in the classrooms. Kshetrimayum (2007) elaborated on multiple perspectives germane to incorporation of ICT in teaching and learning processes: simulation, visualization and modelling constituted the pedagogical perspective; assessment forms cognition perspective; e-learning and virtual learning environments comprise content delivery perspective; and finally, from project work and task perspective ICT with a combination of pedagogy & software design could lead to a collaborative environment.

Challenges ahead

The Human Resource Development department along with the Department of Information Technology has developed a report on technology in education. The report has identified four issues in integrating technology in education in government schools, namely, ICT infrastructure, quality content that is locally relevant, teachers training, and education delivery through public-private partnerships. These four interdependent issues need to be addressed if technology has to be integrated to formal education to improve the quality of education in government schools.

A typical rural school has a number of inherent limitations viz., limited number of qualified teachers, archaic classrooms, chalk-and-talk methodology, variations in curricula (state to state and board to board), unreliable power supply and lack of interactions with the rest of the world .

The advent of cyber cafes and CD-based courseware has opened up possibilities of alleviating the problem. But, the problem still persists i.e., the non-availability of curriculum- and language-specific material backed up by suitably-trained teachers.

The path way for opportunities

The activities use simple and use commonly available software such as Microsoft Power Point. They also provide sufficient flexibility for teachers to modify them according to their objectives, class or pupils' abilities. Much consideration went in the designing of the software such that each lesson constitutes of only a few slides. It was essential that the pupils are not intimidated by the software and hence simple navigation tools were used. We decided against using more advanced authoring software like Mediator, Macromedia Flash or Multimedia Builder as we want Tamil language teachers to modify and enhance our models.

Basic Requirements in using our Microsoft PowerPoint models:

1) Pupils have

- basic IT knowledge
- basic knowledge of Microsoft PowerPoint
- Usage of recording and playback features in PowerPoint
- Typing text and annotating in different colours and sizes

2) Teachers have

- comfortable knowledge in PowerPoint
- using PowerPoint as an effective tool to translate their language teaching strategies into using IT

The design and use of the software took into consideration the following:

- 1) Prior IT knowledge of pupils:
- 2) Time needed to complete the lesson.
- 3) IT equipment generally found in Singapore schools.
- 4) Various skills required for the teacher to design the software

Human Factors experts have researched the strengths and limitations of different types of the physically challenged and come out with norms and standards for equipment design and HCI interfaces. Experts and researchers have exploited the following methods to make the physically challenged as independent as possible:

- a) Audio cues to help the blind
- b) Rich graphics to help the deaf
- c) Combination of the visual and audio cues to help the partially deaf/blind
- d) Flexible input and output devices for the users with limited movements of the limbs and the body
- e) A variety of cues to keep up the interest and excitement in children with learning disabilities

After the stage of memorization and teaching through class representative or leader, blackboard and chalk pieces came about. After which, teaching tools; such as keyboard, computer, smart board, and Tablet PC that consist of computer and mobile phone provide students with the language benefits in class. Tamil letters, Tamil songs, Tamil vegetables, Grandmother stories, are all being sold in the form of CDs/DVDs even in today's commercialized level, and all these have; Tamil's nuances, the beauty of pronouncing in Tamil, vocabulary building in Tamil, India's nature as well as the beautiful Tamil spoken by qualified hosts in their native language that provides a feast for students who hear and view them. Here, the beauty of the language and the benefits of its nativity are displayed in a manner that students can know about. In this stage, we shall see how information technology is used in teaching and learning, at National Institute of Education that trains teachers, who teach Tamil.

Scrimshaw (2004) also points out that the implementation of ICT in the classroom is "both an innovation in technology and teaching" (p.9). On the other hand, multimedia is a combination features of text, graphic, art, sound, animation and video elements with facilities for interaction. Thus, multimedia is a powerful presentation tool, which can be effectively used for teaching. Studies showed that if students are stimulated with audio, they will have about 20 % retention rate, audio-visual is up to 30 % and in interactive multimedia presentation, the retention rate is up to 60% (Vaughan, 1997; p10). Hence, multimedia tools can enhance many skills such as, functional communication as a result of enriched vocabulary, critical and creative thinking.

Government Initiatives and pathway

A need to develop IT Tools to facilitate human-machine interaction, and to promote the use of these tools for various Indian Languages was felt. Towards this, the Department of Electronics of Government of

India had initiated activities in the area of Technology Development for Indian Languages with the following objectives:

- To develop information processing tools to facilitate human machine interaction, information processing in Indian languages and development of multi lingual knowledge systems
- To promote the use of information processing tools for language studies and research.
- To support R&D efforts in the area of information processing in Indian languages covering machine translation, human machine interaction, language learning and natural language processing.

“ICT@school” was one of the important central government schemes formulated in the tenth five-year plan. Its objective was: “promoting usage of ICT in government and government aided schools (particularly in rural areas) and providing ICT infrastructure.” This scheme had four components (GOI, 2007a):

1. Partnership with state governments and union territories for providing computer education and computer-aided education to govt. and govt. aided schools
2. Establishment of SMART schools - which will be technology demonstrators
3. Universalization of computer literacy thru Kendriya Vidyalayas and Navodaya

Vidyalayas and neighbouring schools

4. State Institutes of Educational Training to provided educational content in the form of films, videos and audios etc.

The National Mission for Education through ICT has been envisaged as a federally sponsored scheme to leverage ICT so as to democratize education by providing high quality personalized and interactive knowledge modules over the internet/intranet for all the learners in the higher education institutions. While the eleventh five-year plan (2007-12) has taken into cognizance the importance of education in nation-building, it provided the much needed fillip for expanding the scope of ICT usage in Indian schools. A substantial budget has been allocated towards upgrading technology infrastructure (including ICT) in schools. Targets have been established for reaching out to Upper Primary Schools to have coverage of ICT by 2011-12.

The ICTACT (ICT Academy of Tamilnadu), a consortium of the Centre, State and Confederation of Indian Industry, will provide industry relevant training programmes in the ICT spectrum to the faculty members of the university and its affiliated colleges. Various training programmes will be sponsored by the ICT industry.

End Notes

ICT can help bridge the rural-urban divide, as has been demonstrated by projects that have been successfully rolled out across the globe. Most teachers do utilise programs such as PowerPoint but only as a presentation facility in their IT classes (as a cognitive tool). Apart from this, commercially available software including e-learning platforms, emphasis lies in self-paced learning. Hence, by using IT as an effective tool to develop the metacognitive skills, we hope to improve their oral and aural performance of Tamil Language pupils.

References

1. ^ ^a ^b ^c Gordon, Raymond G., Jr. (ed.), 2005. *Ethnologue: Languages of the World*, Fifteenth edition. Dallas, Tex.: SIL International.
2. ^ Zvelebil 1992, p. 12: "...the most acceptable periodisation which has so far been suggested for the development of Tamil writing seems to me to be that of A Chidambaranatha Chettiar (1907-1967): 1. Sangam Literature - 200BC to AD 200; 2. Post Sangam literature - AD 200 - AD 600; 3. Early Medieval literature - AD 600 to AD 1200; 4. Later Medieval literature - AD 1200 to AD 1800; 5. Pre-Modern literature - AD 1800 to 1900"
3. ^ "Tamil Brahmi script in Egypt". *The Hindu*. 2007-11-21. <http://www.hinduonnet.com/2007/11/21/stories/2007112158412400.htm>. Retrieved 2008-11-11.
4. ^ *India 2001: A Reference Annual 2001*. Compiled and edited by Research, Reference and Training Division, Publications Division, New Delhi: Government of India, Ministry of Information and Broadcasting.
5. Seetha Lakshmi and Jarina Peer.(2009). Use of Tamil Language and IT in Tamil Language Education Redesigning Pedagogy International Conference, National Institute of Education, Singapore
6. Sivagouri S. (2001). Using IT to improve the mother Tongue pupils' oral performance by developing their metacognitive skills. - An Action Research. Research Paper presented at the 5th World Tamil Teachers Conference, held at Singapore on Sept 6-8 2001, and published in the conference proceedings. (pp.188-193)

Enhancing Activity based Tamil Teaching and Learning using Online Video Repositories

A Data Mining based Approach

Dr. K. Vivekanandan
Associate Professor
School of Management
Bharathiar University
Coimbatore - 641 046
vivekbsmed@gmail.com

Dr. V. Saravanan
Professor & Director
Dept. of Comp. Applications
Dr. N.G.P Institute of Technology
Coimbatore - 641 048
profsaran@hotmail.com

Mr. P. Ranjit Jeba Thangaiah
Assistant Professor (SG)
Dept. of Comp. Applications
Karunya University
Coimbatore - 641 114
prjt@live.com

Abstract

Online Teaching and Learning shows a considerable impact in the education system. The activity based education system introduced by the Government of Tamilnadu in the schools shows a remarkable impact on improving student learning across Tamilnadu. This has been witnessed by many visitors across the world wanting to know about this system and the efforts taken by the Government for successful implementation. The video based interactive tutorial helps the students to effectively learn the contents and also interact with the expert. With the decrease in hardware costs every year, storing much video contents is no longer a costlier process. Data Mining techniques and algorithms are the actual tools that analysts have at their disposal to find similar patterns and correlation in the data.

This paper proposes of presenting the activity based Tamil teaching as short video clips. This video clips are available online to all the students. To start with, set of Tamil teachers are chosen as experts for all the Tamil topics of various classes. The experts' lectures/demo is video recorded and stored in the repository. As the lengthiest lectures will reduce the student interest on the subject, all the lecture topics are divided as short videos and made available for students access. The short video lecture will runs from 3 to 5 minutes. When a student browse through a video lecture and finishes the execution, all the other relevant lectures related to the topic viewed by the students also displayed. The students can easily view the other topics without starting searching for another topics or a topic which cannot be searched very easily. Data mining techniques are used to find the identical patterns. These techniques capture the students' behaviors and also trace through the similar navigations performed by other students across Internet. The related short video contents are grouped together and presented to the students. Grouping of similar patterns are carried out by applying data mining techniques such as Association rule mining (determine implication rules for a subset of video lecture attributes, Classification (assign each video record of a database to one of a predefined set of classes analysis) and Clustering Techniques (find groups of similar video records that are close according to some user defined metrics).

This model enhances the activity based learning and surely will create an interest on "Tamil" or Tamil way teaching among all school/college students. This model can also be used to teach other subjects in all

schools/colleges through Tamil. As the telecommunication sector witnesses a major breakthrough using 3G Technologies, all these video lectures can be easily accessed through Mobile phone environment also. The developed model can also be easily implemented by setting up a Multimedia Kiosk/Centre.

Keywords: Data Mining, Teaching & Learning

Introduction

The schools and Institutions of higher education have increasingly embraced online education, and the number of students enrolled in distance programs is rapidly rising in colleges and universities throughout Tamilnadu. In response to these changes in enrollment demands, many schools/colleges/universities have been working on strategic plans to implement online education. The activity based education system introduced by the Government of Tamilnadu in the schools shows a remarkable impact on improving student learning across Tamilnadu. This has been witnessed by many visitors across the world wanting to know about this system and the efforts taken by the Government for successful implementation

Review of Literature

We began writing this paper with a review of past studies of the issues and trends in online teaching and learning in higher education. A recent survey of higher education in the United States reported that more than 2.35 million students enrolled in online courses in fall 2004 [4]. This report also noted that online education is becoming an important long-term strategy for many postsecondary institutions. Given the rapid growth of online education and its importance for postsecondary institutions, it is imperative that institutions of higher education provide quality online programs.

Faculty training and support is another critical component of quality online education. Many researchers posit that instructors play a different role from that of traditional classroom instructors when they teach online courses.

Using Short Video for Lecture/Demo

The video based interactive tutorial helps the students to effectively learn the contents and also interact with the expert. With the decrease in hardware costs every year, storing much video contents is no longer a costlier process.

The following are the few advantages for the students by using the proposed approach

- a. It generates enthusiasm in the students by using interesting or unusual examples.
- b. More useful feedback from students are received by phrasing questions in a positive manner
- c. It engages the students with leading questions to actively involve them in the classroom.
- d. It encourages collaboration among students.
- e. It gives useful comments on homework.
- f. Simultaneously challenge the students to think and also obtain critical feedback about what the students know by asking the right kinds of questions.
- g. Best to explain specific Tamil topics such as Ilakanam, Seiyul and other advanced topics.

This paper proposes of presenting the activity based Tamil teaching as short video clips. This video clips are available online to all the students. To start with, set of Tamil teachers are chosen as experts for all the Tamil topics of various classes. The experts' lectures/demo is video recorded and stored in the repository. As the lengthiest lectures will reduce the student interest on the subject, all the lecture topics are divided as short videos and made available for students access. The short video lecture will runs from 3 to 5 minutes. When a student browse through a video lecture and finishes the execution, all the other relevant lectures related to the topic viewed by the students also displayed. The students can easily view the other topics without starting searching for another topics or a topic which cannot be searched very easily

Using Data Mining Techniques to Enhance Teaching and Learning

Data Mining is the technique to explore and analyze the large data sets, in order to discover meaningful patterns and rules. The evaluation of data mining techniques began when the business data are stored in the database and the technologies were generated to allow the user to navigate the data in the real time. Data mining techniques are used to find the identical patterns. These techniques capture the students' behaviors and also trace through the similar navigations performed by other students across Internet. By considering the proposed objective, this paper proposes the use of data mining techniques to enhance teaching and learning. The major data mining techniques considered in this paper are

- a. Association Techniques.
- b. Classification techniques.
- c. Clustering Techniques.

Association: It is method for discovering interesting relations between the variables in the large database. There are different types of algorithm for association rule. They are Apriori algorithm, éclat algorithm, FP-growth algorithm, One-attribute-rule algorithm, Opus search algorithms, and Zero-attribute-rule algorithm.

Grouping of similar patterns are carried out by applying data mining techniques such as Association rule mining (determine implication rules for a subset of video lecture attributes). The students studying a particular video lecture also studies a related content. The associations between the related contents are displayed at the end of a lecture which helps the student to navigate among the interested pages. This increases the students' interest and also creates an indirect interest on the content and learning.

Using association rule algorithms, the subjects associated with other subjects can be sorted out and an impact can be given to the user that; if the user is viewing/studying a particular subject, the association rule algorithms in turn advises to study/view a video that are related to the subject. In this way the related videos are ranked and shown to the user. This will enhance a user to learn more on a subject with all the associated materials related to the subject.

Classification: It is one of the data mining techniques used to predict the group for data instance. Some of the popular classification techniques are decision trees and neural networks. From the existing database, the end user can classify the land with required parameters like state wise, of district wise, area wise and etc by means of tree like structure.

The above listed classification techniques (assign each video record of a database to one of a predefined set of classes analysis) helps the students to identify a particular subjects and search through specific topics. As the classification process proceeds, the left part or the right part of the tree is considered for better visualization and reading of a subject.

Clustering: It defined as collection of data object that are similar to one another within the same cluster and dissimilar to the objects in the other cluster. Clustering algorithms are broadly classified into hierarchical and partitioning clustering algorithm. Again, the Hierarchical algorithm are Agglomerative and Divisive algorithm and the Partitioning Algorithms are k-means, k-mediod, DBSCAN, CLARA, CLARANS, BIRCH CLIQUE, OPTICS etc.,

Clustering Techniques (find groups of similar video records that are close according to some user defined metrics) groups the related subjects and helps the student in choosing related content without wasting much time of choosing what contents to be chosen next or in future.

Discussion and Conclusion

As schools and institutions of higher education continue to embrace and debate Tamil online learning, it is important to envision where the field is headed. It is the appropriate time for us to propose suitable technical model / approached which will help the Government of Tamilnadu to implement the projects better. This model can also be used to teach other subjects in all schools/colleges through Tamil. As the telecommunication sector witnesses a major breakthrough using 3G Technologies, all these video lectures can be easily accessed through Mobile phone environment also. The developed model can also be easily implemented by setting up a Multimedia Kiosk/Centre.

References

1. C. J. Bonk, "Online Teaching in an Online World" (executive summary), USDLA Journal, Vol. 16, No. 1, January 2002, <http://www.usdla.org/html/journal/JAN02_Issue/article02.html> (accessed August 8, 2006); and C. J. Bonk, "Online Training in an Online World" (executive summary), USDLA Journal, Vol. 16, No. 3, March 2002, <http://www.usdla.org/html/journal/MAR02_Issue/article02.html> (accessed August 8, 2006).
2. C. J. Bonk, *The Perfect E-Storm: Emerging Technologies, Enhanced Pedagogy, Enormous Learner Demand, and Erased Budgets* (London: The Observatory on Borderless Higher Education, 2004); and K.-J. Kim, C. J. Bonk, and T. Zeng, "Surveying the Future of Workplace E-Learning: The Rise of Blending, Interactivity, and Authentic Learning," *E-Learn Magazine*, June 2005, <<http://www.elearnmag.org/subpage.cfm?section=research&article=5-1>> (accessed August 8, 2006).
3. R. Detweiler, "At Last, We Can Replace the Lecture," *Chronicle of Higher Education*, July 9, 2004, p. B8; and R. Zemsky and W. F. Massy, "Why the E-Learning Boom Went Bust," *Chronicle of Higher Education*, July 9, 2004, p. B6.
4. E. I. Allen and J. Seaman, *Growing by Degrees: Online Education in the United States, 2005* (Needham, Mass.: The Sloan Consortium, 2005).
5. <http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/TheFutureofOnlineTeachingandLe/157426>, April 2010.

The Computer, Tamil Language , The Tamil Teacher

Jayasarasvathi DuraiKannu

Bedok West Pri Sch – Singapore

Email: jayasarasvathi_durai_kannu@moe.edu.sg

அறிமுகம்

‘வீழ்வது நாமாக இருப்பினும்’

வாழ்வது தமிழாக இருக்கட்டும்’

என்ற கலைஞரின் வாசகத்திற்கு ஏற்ப ‘காற்று அடைத்த பையான இத்தேகம்’ இன்று தோன்றி நாளை மறையலாம். ஆனால் கல் தோன்றி மண் தோன்றா முன்னே தோன்றிய மூத்தக் குடியின் மொழியானது, தமிழ் மொழி. செம்மொழி என்ற மகுடம் சூடப்பட்டு வீறுநடை போடுகிறது. தமிழ் மொழி பழமையானது என்பது உண்மை. பழமை நிறைந்த அவள் இளமையானவளும் கூட. அவள் இலக்கியங்களில் மட்டும் களிநடனம் புரிவதில்லை, இணையத்திலும் உலா வருகிறாள் என்பது மாணவ சமுதானத்திற்கும் நம் இணைய சமுதாயத்திற்கும் உணர்த்தப்பட வேண்டும். இத்திருப் பணியைத் திறம்பட செய்யக்கூடியவர்கள் ஆசிரியப் பெருந்தகைகள்.

கற்றல் கற்பித்தல் என்பது கல்வி கற்றலின் இரு கூறுகள். ஒரு கூறு சிறப்பாக அமையாவிடின் மற்றொரு கூறு திறம்பட அமையாது.

கணினி மற்றும் இணையம் கற்பித்தலுக்குத் தேவையா?

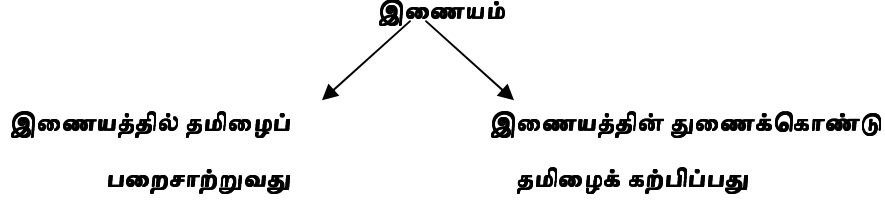
இன்றைய காலத்தில் இணையம் என்பது கல்வி கற்பித்தலில் ஒரு உத்தியாகவே கருதப்படுகிறது. காலம் மாறுது, கருத்தும் மாறுது, நாமும் மாற வேண்டும் என்ற வரிகளுக்கு ஏற்ப சிறார் முதல் பெரியவர் வரை இணையம்பால் கொண்டிருக்கும் ஈடுபாடினைக் கருத்தில் கொண்டு இணையத்தைக் கற்பித்தலுக்கான மேடையாய் அமைத்தல் சால சிறந்தது.

இணையம் வழி தமிழ் கற்பிப்பதால் தமிழ், அழகுத்தமிழாய், பழகு தமிழாய், வளர்தமிழாய், இன்பத்தமிழாய், தீந்தமிழாய், நடைமுறைத் தமிழாய், உருவெடுப்பதைக் கண்கூடாகப் பார்க்கலாம்.

இன்றைய நவீன உலகில் கல்வி கற்பிக்கும் உத்திகளில் கணினி மற்றும் இணையம் முக்கிய இடத்தை வகிக்கிறது. இணையம் என்பது கற்றலுக்கான வழிமுறை மட்டுமில்லை. ஆசிரியர் மாணவர்களிடையே பரஸ்பரம் ஏற்படுத்த வல்லது கணினி/இணையம். இதுவே ‘Speaking the same language’ என்று வர்ணிக்கப்படுகிறது. காலத்திற்கு ஏற்ற கோலம் என்பதற்கு ஏற்ப ஆசிரியர் தன்னை மாற்றுக்கொள்வதும் புதியனவற்றைக் கற்பதும் முக்கியமானது.

வழக்கமான தமிழ் வகுப்பில் தமிழும் இணையமும் சாதாரண தமிழ் ஆசிரியரும்

இணையம் என்று நோக்கையில் அது இரண்டு வகைப்படும். ஒன்று இணையத்தில் மேன் மேலும் தமிழை பறைசாற்றுவது. Blog, Web Page போன்றவற்றில் தமிழில் படைப்புகளை அதிகரிப்பது. மற்றொன்று இணையம் துணைக்கொண்டு தமிழ் மொழியினைக் கற்பிப்பது.



இன்று நான் சராசரி தொழில்நுட்ப திறன் உடைய ஒரு ஆசிரியர் ஒரு சாதாரண தமிழ் வகுப்பில் கணினி மற்றும் இணையத்தைத் தமிழ் கற்பித்தலுக்கு எப்படி எல்லாம் பயன்படுத்தலாம் என்பதை விளக்க முனைகிறேன்.

இத்தருணத்தில் நான் ஒன்றைச் சொல்லிக் கொள்ள விரும்புகிறேன்.

நல்ல ஆசிரியராய் பணி ஆற்றும் பேறு கிடைக்கப்பெறுவது அரிது.

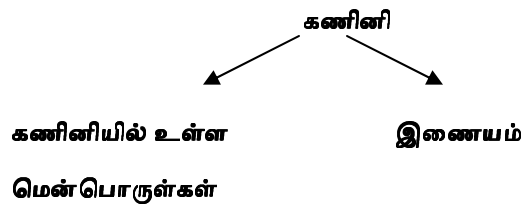
அதனினும் நற்றமிழ் ஆசிரியராய் பணி புரியும் பேறு கிடைக்கப்பெறுவது அரிதினும் அரிது.

இத்தகைய தமிழ் ஆசிரியர்கள், தமிழ் மொழியையும், இலக்கணத்தையும் இலக்கியத்தையும் மட்டும் கற்பிப்பதில்லை. தமிழ் உணர்வையும் தமிழ் ஈடுபாட்டையும் தமிழ் ஆர்வத்தையும் தமிழ் நேசத்தையும் முக்கியமாக தமிழன் என்ற உன்னதமான நேர்த்தியான உணர்வை ஏற்படுத்துவதும் ஒரு தமிழ் ஆசிரியரது கடமையாகும். தமிழ் உணர்வு தமிழ் வகுப்போடு முற்றுப்புள்ளி காணாது தமிழன் இருக்கும் அந்த நாள் வரை தமிழ் உணர்வு அவன் மனதை முற்றுகையிட வேண்டும்.

*தமிழா தமிழா என அழைக்கும்
 ஓசை கேளவில்லையா உனக்கும்
 கேளாதது போல நடிக்கும்
 உன்னைக் கண்டு உலகமே வியக்கும்
 தாய்மொழியானது நடைபோடும் தெய்வம்
 என்பது இல்லையே புதிய தத்துவம்
 இதை மறந்துவிட்ட வர்க்கம்
 நமது என்கிறதே ஊர்சனம்*

போன்ற அவலக் கவிதைகள் உருவாகமல் இருக்க தமிழ் ஆசிரியர்கள்தான் வழி செய்ய வேண்டும். பாடப்புத்தகங்கள், மனனம் - ஒப்பித்தல், பழைய புராணம் பேசுதல், கனி மொழியான தமிழைக் கடும் மொழியாக தோன்ற செய்யும் நடை ஆகியற்றைத் தவிர்த்து புதியனவற்றைப் புகுத்தலில் ஆசிரியர்கள் கண்ணும் கருத்துமாக இருக்க வேண்டும்.

அவ்வகையில் தொழில்நுட்பத்தை இரு கூறுகளாகக் காணலாம்.



மொழிகற்பித்தலில் நான்கு திறன்கள் உள்ளன. அவை பேசுதல், கேட்டல், வாசித்தல். எழுதுதல். இத்திறன்களை மேம்படுத்தும் வண்ணம் நடவடிக்கைகள். தமிழ் சரிவர பேச இயலாத ஒரு சிறுவன், நான் அடிச்சா தாங்கமாட்டே. நாலு மாசம் தூங்க மாட்டே, மோதிப்பாரு வீடு போய் சேர மாட்டேன் என்று உச்சரிப்பு பிறனாது சுருதி கெடாமல் பாடுகிறான். ஆனால் எனது பெயர் இன்னது என்று தமிழில் கூற இயலவில்லை. இதற்குக் காரணம்? ஈடுபாடு, எளிமை, இனிமை. தமிழை ஒவ்வொருவரின் வாழ்நாள் மொழியாக செய்ய வேண்டும் எனில் latest technologies and teaching strategies உடன் தமிழ் தக்க வேகத்தில் செயல்படும்.

முதலில் பெரும்பாலான கணினியில் அமைக்கப்பெற்ற MS Office யின் துணைக்கொண்டு கற்பித்தலை நோக்கலாம்.

முதலில், PowerPoint, என்பதனைக் காணலாம். PowerPoint துணைக்கொண்டு சிந்தையைக் கவரும் படச்சுருள்களை (Slides) உருவாக்கலாம்.

(எடுத்துக்காட்டு - தோசையா - பூசையா?) Powerpoint மூலம் மாணவர்களுக்குக் படங்களுடன் கதைகள் கூறலாம். கதைகள் கூறுவதுடன் ஊகிப்பது, கூர்ந்து கேட்டல் போன்ற திறன்கள் மேம்படுத்தப்படும். மேலும் திறன் குறைந்த மாணவர்கள் கண்டு சயமாகக் கற்று அல்லது மதிப்பீடு செய்யும் வகையில் Powerpoint உருவாக்கலாம். (எடுத்துக்காட்டு காட்டப்படும்)

PowerPoint உதவியுடன் E புத்தகம் உருவாக்கலாம்.

(எடுத்துக்காட்டு - தம்பிப் பாப்பா)

பல வகைகளில் இதனைப் பயன்படுத்தலாம். படங்களைக் காட்டி மாணவர்களுடன் உரையாடலாம். பின்னர் வாக்கியங்களைப் படித்துக்காட்டலாம். பல தொடர் நடவடிக்கைகள் செய்யலாம். **(எடுத்துக்காட்டு)**

Movie Maker போன்ற மென்பொருளால் மாணவர்களது வாய் மொழி ஆற்றலைப் பெருக்கலாம். மாணவர்களை புகைப்படங்களைக் கொண்டு வரச் செய்து அதனை ஒட்டி பேசி பதிவு செய்யச் சொல்லாம்.

(எடுத்துக்காட்டு காட்டப்படும்)

இணையமும் கற்றலும்

இணையம் வழங்கும் தகவல்கள் ஏராளம், அவற்றை வகை கண்டு தேர்ந்தெடுப்பதே பெரும் பணி. U tube இல் காணப்படும் பல படக் காட்சிகள், பாடல்கள், விளம்பரங்கள், வித்தியாசமான தகவல்களின் தொகுப்புகள் கற்பித்தலுக்கு முதுகெலும்பாய் இருக்கிறது. (எடுத்துக்காட்டு காட்டப்படும்) U tube downloaderஐ கொண்டு தரமான நிகழ்ச்சிகளை download செய்துக்கொள்ளலாம். அதுமிட்டும் மின்றி பல தமிழ் இணையத்தளங்களைக் கொண்டு கட்டுரை போன்றவற்றைக் கற்றுக் கொடுக்கலாம்.

(எடுத்துக்காட்டு காட்டப்படும்)

முடிவு

கணினி என்பது இன்றைய வாழ்க்கை முறையோடு பின்னி பிணைந்த ஒன்று. சிறார்கள் முதல் பெரியவர் வரை இதற்கு விதிவிலக்கல்ல. இணையம் வழி தமிழை வளர்ப்போம்! தமிழரின் நிலையை உணர்த்துவோம்!

எம்மொழியாம் செம்மை குன்றா செம்மொழி

இலக்கியம் முதல் இணையம் கண்ட தனிமொழி

பிறவிக்குப் பொருள் தந்த என் தாய்த்திருமொழி

உன்னை வணங்க கண்ணீர் வடிக்கின்றதே என் விழி.

தமிழ் கற்பித்தலில் கணினியும் இணைய வளங்களும்

- என் அனுபவம்

கு. உத்தமன் M.A, M.Phil, Med.

மூத்த தமிழாசிரியர்,

ஆண்டர்சன் உயர்நிலைப் பள்ளி, சிங்கப்பூர்.

kumbalingam_uthamam@moe.edu.sg, / uthamank@hotmail.com

தமிழ்க்கல்வியின் நோக்கம்

சிங்கப்பூரில் தமிழ்மொழி கற்பிப்பது கருத்துப் பரிமாற்றத்துக்கான மொழித்திறன்களைப் பெறச் செய்வதையும் இந்தியப் பண்பாட்டை அறிந்துகொள்ளச் செய்வதையும் நோக்கங்களாகக் கொண்டது ஆகும். எனவே, கேட்டல், பேசுதல், வாசித்தல், எழுதுதல் ஆகிய மொழித்திறன்களை வளர்க்கும் வகையில் சிங்கப்பூர் மாணவர்களுக்கென பயிற்றுக்கருவிகள் கல்வி அமைச்சின் பாடத்திட்ட வரைவு மற்றும் மேம்பாட்டுப் பிரிவால் உருவாக்கப்படுகின்றன. ஆனபோதிலும் அவற்றைப் பயன்படுத்திக்கொள்வதில் நீக்குப்போக்கைக் கடைப்பிடிக்க ஆசிரியர்களுக்கு உரிமை உண்டு. அதாவது, ஆசிரியர்கள் தம் மாணவரது திறனையும் தேவையையும் கருத்திற்கொண்டு பாடநூல்களுக்கு வெளியிலிருந்து கற்பித்தலுக்கான வளங்களைத் தேடித் திரட்டிப் பயிற்றுவிக்கலாம். அதுமட்டுமல்லாமல், சிங்கப்பூர் பள்ளிகள் கற்றல் கற்பித்தலில் புதுமையைப் புகுத்த ஊக்குவிக்கப்படுகின்றன. இந்த அடிப்படையில்தான் அவ்வப்போது இணையத்தில் உள்ள வளங்களைத் தமிழ்க் கற்பிக்க நான் பயன்படுத்திவந்துள்ளேன். அவை பயன்படுத்தப்பட்ட வழிவகைகளை இங்கு விளக்க முற்படுகிறேன்.

ஒலி ஒளிக் காட்சிகளைப் பயன்படுத்திக் கேட்டலும் பேசுதலும்

சிங்கப்பூரில் உயர்நிலைப் பள்ளிகளில் பயிலும் மாணவர்களை அவர்களிடத்துள்ள கற்கும் ஆற்றலின் அடிப்படையில், வழக்கத்தொழில் நுட்பம், வழக்கம், விரைவு எனத் தர வரிசைப்படுத்தலாம். இவர்கள் முறையே அடிப்படைத் தமிழ், தமிழ், உயர்தமிழ் என்னும் பிரிவுகளில் தமிழ்மொழியைக் கற்பார்கள். வழக்கத்தொழில் நுட்பப் பிரிவில், அடிப்படைத் தமிழ் கற்கும் மாணவர்கள் பெரும்பாலும் கேள்விக் காட்சி வழிக் கற்கும் ஆற்றல் உள்ளவர்களாவும் செயல்வழிக் கற்பதில் நாட்டம் உள்ளவர்களாவும் இருப்பார்கள். இதைக் கருத்திற்கொண்டு இப்பிரிவில் முதலாமாண்டு பயிலும் மாணவர்களுக்குத் தமிழ் கற்பிக்க (13 வயதினர்) <http://www.youtube.com> என்னும் இணையத்தளத்தில் உள்ள முல்லாக் கதைகள் சிலவற்றைப் பயன்படுத்திக்கொண்டேன். மாணவர்கள் ஒலி ஒளியுடன் கூடிய அக்கதைகளைக் கணினியில் பார்த்தும் கேட்டும் மகிழ்ந்தனர். அக்கதைகள் வலியுறுத்தும் நற்பண்புகள் குறித்துக் கலந்துரையாடினர். "குழந்தையின் கேட்டல் அளவினைப் பொருத்தே மொழியைப் பேசும் அளவும் திறனும் அமைகிறது." ¹

சிறுவர் பக்கத்தின் துணைகொண்டு வாசித்தலும் நடித்தலும்

இதுதவிர, <http://www.moderntamilworld.com>, என்னும் இணையத் தளத்தில், சிறுவர் பக்கம் என்னும் பகுதியில் இடம்பெற்றுள்ள சின்னஞ்சிறு கதைகள், பாடல்கள் இம்மாணவர்களின் தரத்துக்கு ஏற்புடையனவாக இருந்தன. அதனால், அவற்றுள் சிலவற்றை வாசிக்கத் தூண்டினேன். வாசிப்புக்குப்

¹ "குழந்தையின் கேட்டல் அளவினைப் பொருத்தே மொழியைப் பேசும் அளவும் திறனும் அமைகிறது."

பிந்திய நடவடிக்கையாக மாணவர்கள் கதையை நடித்துக்காட்டவும் கதை குறித்துக் கலந்துரையாடவும் ஊக்குவிக்கப்பட்டனர். மாணவர்களின் ஈடுபாடு விரும்பத் தக்கதாக இருந்தது. மேலும், இன்று கல்வியில் பன்ம நுண்ணறிவுத் திறன்கள் (Multiple Intelligences) வேறுபடுத்தப்பட்ட கற்பித்தல் (Differentiated Instruction), மாணவர்களின் கற்றல் பாணி (Learning Style) முதலிய அணுகுமுறைகள் பரவலாகப் பேசப்பட்டும் பின்பற்றப்பட்டும் வருவதும் இங்கே கருதிப்பார்க்கத்தக்கது.

வாசித்தலும் வாசிப்பை இணையத்தில் ஏற்றம் செய்தலும்

<http://www.voivethread.com> என்னும் இணையத்தளம் மாணவரின் உச்சரிப்பு, சொல்லழுத்தம் ஆகியவற்றை வளர்க்க உதவியாக உள்ளது. எவ்வாறெனில், இரண்டொரு பத்தியில் அமைந்த பனுவலை (ஏறக்குறைய 120 சொற்களில் அமைந்தது) PDF கோப்பாக ஏற்றம் செய்து வைத்திட வேண்டும். அதோடு, அப்பகுதியை ஆசிரியரே ஒருமுறை வாசித்து மாதிரி வாசிப்புக்காக (Model Reading) ஒலிவடிவில் அங்கேயே ஏற்றம் செய்துவைத்திடவும் வேண்டும். குரலைப் பதிவு செய்வதற்கான வசதி அவ்விணையத் தளத்தில் உண்டு. பிறகு, ஆசிரியர் ஏற்றம் செய்து வைத்துள்ள வாசிப்புப் பனுவலை மாணவர்கள் படிக்கவும், மாதிரி வாசிப்பைக் கேட்கவும் செய்திட வேண்டும். அடுத்து, மாணவர் பனுவலை வாய்விட்டு வாசித்துத் தமது வாசிப்பைப் பதிவு செய்யவும், பதிவு செய்த தமது வாசிப்பை இணையத்தில் ஏற்றமும் செய்ய முடியும்.

மேற்கூறியவாறு மாணவரின் வாசிப்பாற்றலை வளர்க்க முயன்றேன். <http://voicethread.com/#u827011.b1040505.i5546639> என்னும் இணையப் பக்கத்திலும், <http://voicethread.com/#u827011.b1044272.i5568261> என்னும் இணையப் பக்கத்திலும் மாணவர்கள் வாசித்துத் தம் குரலைப் பதிவு செய்திருப்பதைக் கேட்கலாம். இக்கற்றல் நடவடிக்கையில் மாணவர்கள் மகிழ்ச்சி ஆரவாரத்துடனும் ஆர்வத்துடனும் ஈடுபட்டனர். சுருங்கச் சொன்னால், தமிழ் ஒலி, கணினிக்கூடமெங்கும் நிறைந்து வழிந்தது எனலாம். இத்தகைய பயிற்சி முன் குறிப்பிட்ட மூன்று பிரிவுகளைச் சேர்ந்த மாணவர்களுக்கும் வழங்கப்பட்டது. ஏனெனில், எல்லாப் பிரிவு மாணவர்களுக்கும் வாசிப்புத் திறனில் தேர்வு உண்டு.

கருத்தறிதல் பயிற்சியும் அகன்ற வாசிப்பும்

வழக்கம், விரைவு ஆகிய பிரிவு மாணவர்கள் ஓரளவு நீண்ட பனுவலை வாசிக்கும் திறன் படைத்தவர்கள். இவர்களிடையே வாசித்துக் கருத்தறியும் திறனை வளர்ப்பது மிகவும் தேவை. ஏனெனில், இவர்களுக்குத் தேர்வில் கருத்தறிதல் திறன் ஒரு முக்கியக் கூறாகும். http://tamil.webdunia.com/miscellaneous/literature/remembrance/0806/15/1080615009_1.htm எனும் இணையப்பக்கம் உலகெங்கும் கொண்டாடப்படும் “தந்தையர் தினம்,” எங்கே, எவ்வாறு, யாரால் உருப்பெற்றது என்பதை வரலாற்று ரீதியில் அழகாகப் படம் பிடித்துக் காட்டுகிறது. எனவே, குடும்ப உறவின் உன்னதத்தை உணர்த்தவும் தந்தையரின் அர்ப்பணிப்பு உணர்வை அறியச் செய்யவும் இப்பகுதியைத் தமிழ்க் கற்பித்தலுக்குப் பயன்படுத்திக்கொண்டேன்.

அடுத்து “அஜந்தா ஓவியங்கள்” குறித்து ஒரு கட்டுரை http://tamil.webdunia.com/entertainment/tourism/tourismspots/0804/30/1080430017_1.htm எனும் இணையப்பக்கத்தில் இடம்பெற்றுள்ளது. தமிழ்மொழி கற்றலின் நோக்கங்களுள் இந்தியர்களின் மரபு சார்ந்த கலை, பண்பாடுகளை அறிந்து போற்றுவதும் ஒன்று என்பதால் அப்பகுதியையும் பயிற்றுவித்தலுக்குப் பயன்படுத்திக்கொண்டேன். மாணவர்கள் முதலில் இணையத்தில் இக்கட்டுரைகளைப் படித்தார்கள். பின்னர், அக்கட்டுரை தொடர்பாக ஆசிரியர் தயாரித்து அளித்த வினாத்தாளில் விடைகளை எழுதி ஒப்படைத்தார்கள். இங்ஙனம் இணையப் பனுவல்கள் வழிக் கருத்தறியும் திறனை வளர்க்கும் நடவடிக்கையை மேற்கொண்டேன். இம் மாணவர்களிடையே அகன்ற வாசிப்பை வளர்க்க. <http://www.thamilworld.com/forum/lofiversion/index.php?t12711.html>.

<http://www.thamilworld.com/forum/lofiversion/index.php?t12551.html> எனும் இணையப் பக்கங்களில் உள்ள கதைகளை வாசிக்கச் செய்து அவை வலியுறுத்தும் நற்பண்புகளை அடையாளங் கண்டு அவற்றை எடுத்து எழுதப் பணித்தேன். 'மாணவரிடையே படித்தலின் அளவு, விரைவு ஆகியவற்றைப் பெருக்கும் வழிகளைக் காண வேண்டும். பாட நூல்களை மட்டுமே படித்தல் பொருளாகக் கருதாமல், இதழ்கள், அறிக்கைகள், துணைநூல்கள் ஆகியவற்றையும் மாணவர்களுக்கு வழங்க வேண்டும்.'² நூலகப் புத்தகத்தின் துணுகொண்டு வாரம் ஒருமுறை அகன்ற வாசிப்பில் மாணவரை ஈடுபடுத்தி வருகிறேன். ஆனால், ஒரு மாற்றத்துக்கு மேற்கூறப்பட்ட இணையக் கதைகள் பயன்படுத்திக்கொள்ளப்பட்டன.

மொழிப்பயிற்சிகளும் மீள் வலியுறுத்துதலும்

மாணவர்கள் வகுப்பில் பயிலும் பாடங்கள் தொடர்பான சிலவற்றை மீள் வலியுறுத்தும் நோக்கில் பல தரப்பட்ட பயிற்சிகளைத் தயாரித்து <http://www.andersonsec.moe.edu.sg/ebook/index.asp?subject=Tamil+Language> என்னும் எங்கள் பள்ளி இணையப் பக்கத்தில் ஏற்றிவைத்துள்ளேன். இந்தப் பயிற்சிகள் யாவும் 'Hot potatoes' என்று சொல்லப்படும் படைப்புக்கருவியைப் (Authoring Tool) பயன்படுத்தித் தயாரிக்கப்பட்டவை. இம்மென்பொருளைப் பயன்படுத்தி மரபுத்தொடர்கள், பெயரடை (குறிப்புப் பெயர்ச்சம்), வினையடை (குறிப்பு வினையெச்சம்), இரட்டைக் கிளவிகள், அடுக்குத் தொடர்கள், வேற்றுமை உருபுகள், வாக்கிய இயைபுகள், விடுகதைகள் முதலான மொழிப் பயிற்சிகளைத் தயாரித்துப் பள்ளி இணையப்பக்கத்தில் இட்டு வைத்துள்ளேன்.

தவிர, பத்தியில் கோடிட்ட இடங்களில் பொருத்தமான சொற்களை நிரப்பும் பயிற்சியும் உண்டு. இதன்வழி இணைப்புச் சொற்கள், மறுப்புச் சொற்கள், அருஞ்சொற்கள் முதலான சொற்களைத் தேர்ந்தெடுத்து நிரப்புவார்கள். மாணவர்கள் மொழிப்பயிற்சிகளைச் செய்யும் அதே நேரத்தில் தகவல்களையும் அறிந்துகொள்ள வேண்டும் என்பதால் பிள்ளைகளின் கடமை, மூப்படையும் மக்கள் தொகை, சுற்றுச் சூழல், சுற்றுலாத்தளம், பரதநாட்டியம், தாய்நாட்டுப் பற்று, பெண்களின் வீரம், மரமும் மழையும் முதலான பலதரப்பட்ட தலைப்புகளின் கீழ் இப்பத்திகள் அமைக்கப்பட்டன. இவை தவிர, திரையிசைப் பாடல்களையும் இவ்வகைப் பயிற்சிக்குப் பயன்படுத்திக் கொண்டிருக்கிறேன். மாணவர்கள் பயிற்சிகளைச் செய்யும்போது தங்களின் விடைகளைத் தாங்களே சரிபார்க்கும் வசதிகள் உண்டு. எனவே, இது இருவழித் தொடர்பிலான பயிற்சிகள் (Interactive exercise) ஆகும். மேலும், மாணவர்கள் இப்பயிற்சிகள் தொடர்பாக தங்களது கருத்துகளைத் (Feedback) தெரிவிக்கவும் இடம் உண்டு. அது நம் பயிற்சிகளின் குறை நிறை குறித்து நாம் அறிந்து அவற்றை மேம்படுத்தவும் மாற்றியமைக்கவும் உதவும்.

இவை தவிர, எம் பள்ளியில் ஆண்டுதோறும் குறிப்பிட்ட பிரிவில் பயிலும் மாணவர்கள் ஓரிரு நாட்களுக்கு இணையத்தைப் பயன்படுத்திப் பாடம் செய்ய வேண்டியிருக்கும். இதற்காகப் பள்ளி இணையத்தளத்தில் பாடம் இட்டு வைக்கப்படும். பள்ளிக்கு வராமலேயே மாணவர்கள் தத்தம் இல்லங்களில் கணினிவழி அப்பாடங்களைச் செய்ய வேண்டும். இதை "E learning Day" என்று குறிப்பிடுவதுண்டு. இத்தருணத்தில் மாணவர்களைக் கற்றலில் ஈடுபடுத்த இத்தகைய பயிற்சிகள் உதவியாக இருந்தன.

இணையமும் இலக்கியமும்

உயர்தமிழ் படிக்கும் மாணவர்களுக்கு இலக்கியத்தை அறிமுகப்படுத்தி அவர்களின் தமிழார்வத்தையும் தமிழறிவையும் மேம்படுத்த வேண்டியது அவசியமாகிறது. இவர்களுக்கு எளிய கவிதைகள், உள்ளூர்ச் சிறுகதைகள் ஆகியவற்றை அறிமுகப்படுத்திக் கற்றல் நடவடிக்கைகளில் ஈடுபடுத்துவதுண்டு.

உயர்நிலை மூன்றாம் வகுப்பில் உயர்தமிழ் பயிலும் மாணவர்க்குப் பாரதியார் கவிதைகளை அறிமுகப்படுத்த விரும்பினேன். அதற்காகச் 'சிந்து பைரவி' எனும் திரைப்படத்தில் நடிகை சுஹாசினி, நடிகர் சிவக்குமாருக்குப் பாரதியார் கவிதைகளை அறிமுகப்படுத்தும் காட்சியையும் அதனைத் தொடர்ந்து நடிகர்

² இரத்தினசாபதி, பி. (2005: பக்க எண் 196). *தமிழ் கற்க கற்பிக்க*, மயில வேலன் வெளியீடுகள், சென்னை.

சிவக்குமார் கடற்கரையில், “மனதில் உறுதி வேண்டும்’ என்னும் பாரதியாரின் கவிதையைப் பாடுவதையும் காண்பித்தேன். அதற்கு அடுத்த கட்டமாக மாணவரிடம் பாரதியாரின் கவிதைகள் குறித்து ஓரளவு கூறினேன். குறிப்பாக, பாரதியாரின் ‘கண்ணன் பாட்டுக் குறித்துச் சொன்னேன். அடுத்து, <http://www.moderntamilworld.com/illakiyam/bharathiyar-kannanpattu8.asp> என்னும் பக்கத்தில் உள்ள “கண்ணன் - என் விளையாட்டுப் பிள்ளை’ என்னும் தலைப்பிலான பாரதியாரின் கவிதையை வாசிக்கச் சொன்னேன். நானும் வாசித்துக் காட்டினேன். அதன் பிறகு, அதன் நயங்களை விளக்கிக் கூறினேன். பின் நடவடிக்கையாக, தயாரித்து வைக்கப்பட்டிருந்த பயிற்சித்தானை வழங்கி அவற்றில் விடை எழுதுமாறு பணித்தேன். மாணவர்கள் தமிழ்ப்பாடத்தை மிகவும் அனுபவித்துக் கற்றனர்.

இதுதவிர, கவிஞர் வைரமுத்து ‘நயாகரா அருவி’ குறித்து எழுதிய கவிதை, சிங்கப்பூர்க் கல்வி அமைச்சு வெளியிட்டுள்ள பாடநூலில் உயர்நிலை நான்காம் வகுப்பில் உயர்தமிழ் மாணவர்க்குப் பாடமாக வைக்கப்பட்டிருக்கிறது. அப்பாடத்தைக் கற்பிக்கையில், கவிஞர் வைரமுத்து அவர்களே அக்கவிதையை உணர்ச்சியோடும் ரசனையோடும் படித்துக் காட்டும் ஒலிக்கோப்பை இணையத்திலிருந்து பதிவிறக்கம் செய்து வகுப்பறையில் ஒலிபரப்பினேன். கவிஞரின் கவிதை ஒலிக்கும் அதேவேளையில், பின்னணியில் நயாகரா அருவிப் படங்களைக் கொண்டு தயாரித்து வைத்திருந்த பவர்பாயண்ட் ஷோவை (Power Point Show) திரையில் தொடர்ந்து ஓடிக்கொண்டிருக்கச் செய்தேன். பிறகு கவிதையின் பொருளை விளக்கிக் கூறினேன். மாணவர்கள் மிகவும் அனுபவித்துக் கற்றனர். “கல்வி சார்ந்த இலக்கியப் பாடங்களும் ஒரு நுகர்பொருளாகவே கருதப்படுகின்றன. எனவே, கற்போரின் தேவையையும் பயன்பாட்டையும் ஆர்வத்தையும் மனத்திற்கொண்டே இலக்கியப் பாடங்கள் தயாரிக்கப்படல் வேண்டும்.”³

மாணவர்களின் படைப்புகளும் படங்களும்

மாணவர்களிடையே எழுதுதல் திறனை வளர்க்கும் நோக்கில் கடிதம், உரையாடல், கட்டுரை ஆகியவற்றை எழுதுவதற்குப் பயிற்சிகள் அளிக்கப்படும். அங்ஙனம் அளிக்கப்பட்டபோது மாணவர்கள் செய்து ஒப்படைத்த தரமான எழுத்துகளை <http://andersontamils.blogspot.com> என்னும் வலைப்பூவில் காணலாம். மூத்த மாணவர்கள் பள்ளியை விட்டுச் செல்லும்போது, புதிதாகப் பள்ளிக்கு வரும் இளைய மாணவர்களுக்கு இப்படைப்புகள் வழிகாட்டுதலாக இருக்கும். என்பதற்காகவே இவ்வாறு செய்யப்பட்டுள்ளது. அங்கே இந்திய கலை, பண்பாடு முதலான சில படங்களையும், மாணவர்கள் போட்டிகளில் வென்று பரிசு பெறும்போது எடுக்கப்பட்ட புகைப்படங்களையும் இட்டு வைத்துள்ளேன்.

நிறைவு

இதுகாறும், கட்டுரையில் இணையத்தை ஒரு வளமை நிலையமாகக் கருதி அதிலுள்ள ஒலி ஒளிக் காட்சிகளைப் பயன்படுத்தித் தமிழ் கற்பித்ததைப் பார்த்தோம். மேலும், இணையக் கட்டுரைகளையும் கதைகளையும் கருத்தறி திறனை வளர்க்கவும் நற்பண்புகளை உணர்த்தவும் அகன்ற வாசிப்புக்கும் பயன்படுத்திக்கொள்ள முடியும் என அறிந்தோம். தவிர, மாதிரி வாசிப்பை ஆசிரியர் உருவாக்கி வைக்கவும், மாணவர் தாமே வாசித்துத் தமது வாசிப்பைப் பதிவு செய்வதற்கு வாய்ப்பு இருப்பதாலும் பள்ளி இணையத்தளத்தில் மொழிப்பயிற்சிகள் ஏற்றம் செய்யப்பட்டுப் பயிற்றுவித்தலுக்குப் பயன்படுத்திக் கொள்ளப்படுவதாலும் இணையத்தைக் கற்றல் கற்பித்தலுக்கு உரிய துணைக்கருவியாகவும் பயன்படுத்திக்கொள்ள முடியும் என்று கண்டோம். இவை இணையத்தைப் பயன்படுத்திக்கொள்ளும் வழிகளுள் சிலதான். இன்னும் பல வழிகளிலும் பயன்படுத்த முடியும்.

³ இரபிசிங், ம.செ. (2005: பக்க எண் 225) *இன்றைய சூழலில் இலக்கியம் கற்பித்தல்*. 5 வது உலகத் தமிழாசிரியர் மாநாட்டு மலர், சிங்கப்பூர்த் தமிழாசிரியர் சங்கம், சிங்கப்பூர்.

வகுப்பறைக் கணினியில் தமிழ்க் கல்வி

N.Vairamani, M.COM., M.C.A.,[M.B.A]

Raja Rajeswari Engineering College,

Chennai -95

"எளிய நடையில் தமிழ் நூல்கள் எழுதிடவும் வேண்டும்
இலக்கண நூல் புதிதாக இயற்றுதலும் வேண்டும்
வெளியுலகில் சிந்தனையில் புதிது புதிதாக
விளைந்துள்ள எவற்றினுக்கும் பெயர்களெல்லாம் கண்டு
தெளிவுறுத்தும் படங்களொடு சுவடியெல்லாம் செய்து
செந்தமிழைச் செழுந்தமிழாய்ச் செய்திடவும் வேண்டும்
உலகியலின் அடங்கலுக்கும் துறைதோறும் நூற்கள்
ஒருத்தர் தயை இல்லாமல் ஊரறியும் தமிழில்
சலசலென எவ்விடத்தும் பாய்ச்சிட வேண்டும். . . "

- பாவேந்தர் பாரதிதாசன்

மொழி என்பது வாழ்வுக்கு ஒளி போன்றது. தமிழில் கற்பதும் கற்பிப்பதும் புதுமையாகவும் எளிமையாகவும் இருக்க முயல்வது அறிவுடைமையாகும். உலக மொழிகள் பலவற்றில் வகைப்படுத்தப்பட்டுள்ள மொழிக்குடும்பங்கள் அனைத்திற்கும் வேர்ச்சொல்லை வழங்கியது தமிழ் மொழி என்பது அறிஞர்கள் ஏற்றுக்கொண்ட கருத்து. உலகில் முதன் முதலில் எழுத்துரு வடிவம் பெற்றது சுமேரிய மொழி என வாதிடுவோர் ஒரு புறமிருக்க எழுத்துரு அடைந்த மொழி தமிழ் என ஆரியரான ராகுல சாங்கிருத்தியாயன் தனது "வால்காவிலிருந்து கங்கை வரை" என்ற நூலில் குறிக்கிறார். ஒரு மொழியைக் கற்க அந்த மொழியின் வேர்களையும், கூறுகளையும் அறிவது அவசியம். வேர் என்பது தோற்ற காலமும் தோன்றிய இடமும். கூறுகள் என்பது அதன் கிளையும், கலையும் ஆகும். அந்த வகையில் தொன்மை மொழியாகிய தமிழில் கற்க இன்று முதலில் வேண்டுவது அதன்பால் ஆழ்ந்த நம்பிக்கை. தமிழைக் கற்பதால் நாம் இந்த உலகில் நிச்சயமாக உயர்ந்த நிலையை அடையலாம். இத்தகைய சிறப்பு வாய்ந்த தமிழ் மொழி கணினி வழிக் கல்வித்துறையில் பல பரிணாமங்களைக் கடந்து விட்டது. தமிழ் கற்றல்,கற்பித்தல் பணியில் மேம்பாட்டிற்கு உதவும் பயன்படு பொருளாகவும் பாடம் பயின்றும் ஆசிரியராகவும் கணினி பயன்படுகிறது. கற்பித்தலில் புதுமையையும் புரட்சியையும் ஏற்படுத்தும் ஆற்றல் கணினிவழிக் கல்விக்கு உண்டு. திட்டமிட்ட பாடப்பொருள் வழியாக கணினிக் கல்வி அளிப்பதால் மாணவரின் கற்றல் நிலை மேம்பட வழி உண்டு.தமிழ் எழுத்துக்களையும், தமிழ் இலக்கியங்களையும், நழுவல் படக்காட்சிகளாகத் (Power Point Presentation) தயாரித்து வகுப்பறையில் கணினி வாயிலாக எளிமையான முறையில் தமிழைக் கற்பிக்கலாம். இதற்குத் தமிழாசிரியர்கள் அடிப்படைத்தகவல் தொடர்புத் தொழில் நுட்பத்திறன் பெற்றிருப்பவராக இருத்தல் வேண்டும். பள்ளியிலும் பள்ளிக்குப் புறத்தேயும் இதர ஆசிரியர்களுடன் இணைந்து எளிய முறையில் பாடத்திட்டத்திற்கேற்ப நழுவல் படக்காட்சிகளைக் குறுந்தகட்டில் (Compact Disc) பதிவு செய்யலாம். பல எழுத்துருக்களைப் (Fonts) பயன்படுத்தாமல் ஒருங் குறியீட்டு (Unicode) முறையினைப் பயன்படுத்தலாம். தமிழுக்குத் தனி அடையாளம் அதன் பண்பாடு. தமிழ் நாட்டில் தமிழ் கற்பது பண்பாட்டிற்காகவும், பொருளாதாரத் தேவைக்காகவும் உள்ளது. ஆனால் தமிழ் பேசும் பிற நாடுகளில் தமிழ் மொழியினைப் பண்பாட்டை மேம்படுத்தும் மொழியாகவே பயன் படுத்துகின்றனர். இதனால், பிற நாடுகளில் தமிழ் கற்பது எளிதான செயலாக வருங்காலத்தில் உருவாகாது. தமிழ் எழுத்துகளைக் கற்பிப்பதில் எளிய வடிவமைப்பு

முறைகளைக் கையாளுதல் வேண்டும். எளிமையாகத் தமிழ் கற்பதற்கு ஏற்றவாறு எழுத்துகள் அமைய வேண்டியதன் அவசியம் இன்றைய காலகட்டத்தில் அதிகரித்துள்ளது. 12 வருடங்கள் தமிழ் மொழியினை முழுமையாகக் கற்றும் தமிழ் மொழியில் படைப்பாற்றல் பெறும் திறனின்றி பிழையின்றி எழுதத் தெரியாதவராய் உள்ளனர்.

இதற்கு

- பலமுறை கற்றுக் கொடுத்துப் பயிற்சி அளிக்காமை
- மொழியில் உள்ள எழுத்துகள், சொற்களின் ஒலிப்புகளைச் சரியாகப் புரிந்து கொண்டு உச்சரிக்க இயலாமை
- சொற்பொருளை அறிந்து படிக்க இயலாமை
- மொழிநடை அமைப்பை, எழுத்துச் சொல் இணைப்பைப் புரிந்து கொள்ளாமை
- தவறாகப் படித்து பிறரின் கேலிக்கு ஆளாகி விடுவோமோ என்ற அச்சஉணர்வு,
- தன்னம்பிக்கை இன்மை போன்றவை காரணங்களாகும்.

பள்ளியில் கணினி வசதியின்மை, தேர்வுக்கு முக்கியத்துவம் அளிக்கும் மனப்போக்கு, தொழில்நுட்ப மாறுதலுக்குத் தங்களைத் தயார்படுத்திக் கொள்ளாமை, வேலைப்பளு, ஆர்வமின்மை போன்றவை தமிழ் கற்பித்தலில் தடைகளாக உள்ளன. பல்லாடகத் தகவல் தொடர்புத் தொழில்நுட்பம் பற்றிய சிந்தனைகள் தமிழ் மொழியின் சிறப்பினை வெளிக்காட்டாது, கணினி வழியாகத் தமிழுக்கே உரித்தான ஏற்ற இறக்கத்துடன் கூடிய ஒலி உச்சரிப்புகளையும், நழுவல் படக்காட்சிகள் வாயிலாகவும் மொழியினைச் சிறப்பாகக் கற்பித்தல் இயலாத செயலாகும், பழைய போதனா முறையே சிறப்பானது என்பன போன்ற தவறான சிந்தனைகள் தமிழாசிரியர்களின் மனதிலிருந்து அகல வேண்டும்.

உலகளவில் ஏழரை கோடி தமிழர்களில் ஒன்றரை கோடி தமிழர்கள் தமிழகத்திற்கு அப்பால் ஐம்பதிற்கும் மேற்பட்ட நாடுகளில் வசிக்கின்றனர். மொரிசியஸ், கரிபியன், அந்தமான், தென் ஆப்ரிக்கா போன்ற தமிழ் மக்கள் வாழும் பல நாடுகளிலும் தமிழ்க் கற்றலுக்கான வசதிகள் இல்லை. அமெரிக்காவில் கலிபோர்னியா தமிழ்க் கல்வி நிலையம், நியூயார்க்கில் உள்ள சிசியா கல்வி மையம், நியூ செர்சி தமிழ்க்கலை மற்றும் பண்பாட்டுச்சங்கம் போன்றவை தமிழ் இணையதளத்தின் (www.tamilvirtualuniversity.com) தமிழ்ப் பாடங்களைப் பின்பற்றுகின்றன. தமிழ் மொழியில் உள்ள 247 உயிர் மெய் எழுத்துகளின் குறியீடுகளை நழுவல் படக்காட்சிகள் மூலம் அளிப்பதால் கற்றல் திறன் பெருகும். தமிழ் மொழியினைத் தெளிவுபடக் கற்க இயலும்.

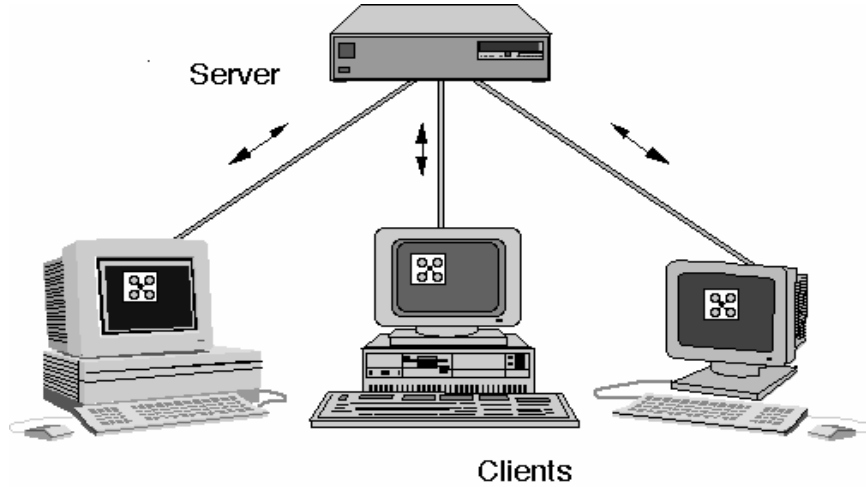
அ	ஆ	இ	ஈ	எ	ஏ	ஐ	ஓ	ஔ	ஊ
-	ஈ	ி	ீ	ெ	ே	ை	ொ	ோ	ௌ
			உ			ஊ			
ு	ி	ு	ூ	ு	ு	ு	ு	ு	ு

இவ்வரைபடத்தில் உள்ள உயிரெழுத்தான (உ,ஊ) விற்கு உயிர் மெய்யெழுத்து வரி வடிவக்குறியீடுகளைச் சரிவரப் பயன்படுத்துதல் வேண்டும். பின்வரும் எழுத்துகளை வரியொற்றி எழுதும்முறையை நழுவல் படக்காட்சிகளாக அளித்தால் எழுத்துக்கள் எழுதும்போது பிழை ஏற்படாது.

உ	----	ங,ச,ப,ய,வு
உ	----	கு,டு,மு,ரு,ழு,ளு
உ	----	னு,ணு,து,நு,லு,று,னு
ஊ	----	கூ,கூ,பூ,பூ,ஊ
ஊ	----	கூ,டு,மு,ரு,ழு,ளு
ஊ	----	கூ,ணூ,தூ,நூ,லூ,றூ,னூ

மொழியறிவு பெறக் கேட்டல் அவசியமாகிறது. பிறர் கூறும் சொற்களைக் கேட்டுணர்ந்து மனதிற்குள் உச்சரிக்கப்படுகிறது. பின்னர் அச்சொற்களின் ஒலி வடிவத்தை வரி வடிவமாகக் கண்களால் பார்த்து வாய் விட்டுக் கூறுகிறது. இதனால் சொற்களுக்கும் அவை குறிக்கும் பொருளுக்கும் உரிய தொடர்பை அறிய முடிகிறது. இச்செயல்முறையில் ஏதேனும் இடர்ப்பாடு ஏற்பட்டால் அது படித்தலைப் பாதிக்கும் என்று அறிஞர் 'முல்லர்' கூறியுள்ளார். இதனால் கேட்டல் திறனை வளர்க்கும் வகையிலும் கணினி வழிக் கற்பித்தல் அமைதல் வேண்டும்.

பல வகுப்பறைகளைக் கொண்ட ஒரு பள்ளியில் ஒரு கணினியை தொடர்புப் பகிர்வி (Communication Server) ஆகவும், மற்ற கணினிகளைச் சார்பு (Clients) களாகவும் நிறுவி பகிர்வியிலிருந்து தகவல்களை எளிய முறையில் பரிமாறிக்கொள்ள முடியும். இதன் வழியாக எளிமையான முறையில் தமிழ்க் கல்விக்கான பாடங்களைப் பகிர்விக் கணினியில் சேமித்து, சார்புக் கணினிகள் வாயிலாக வெவ்வேறு வகுப்பறைகளில் தகவல்களைப் பெற்றுக் கொள்ளலாம்.



"புலன்காட்சியின் வழியே பெறப்படும் அனுபவங்களே கல்வியின் அடிப்படையாகும்" என்பர் அறிஞர் காமினியஸ். கணினியின் ஒளித்திரை (Liquid Crystal Display) மூலம் நழுவுப் படக்காட்சிகளை மாணவர்களுக்கு அளிப்பதன் மூலமும் தமிழ்ப் பாடங்களை எளிதில் புரிய வைக்கலாம்.

எந்த ஒரு நாட்டின் முன்னேற்றத்திற்கும் படைப்பாற்றல் இன்றியமையாததாகும். மாணவர்களின் படைப்பாற்றல் திறனை அதிகரிக்கக் கணினி மென்பொருள் வசதியுடன் கூடிய மொழிப் பயிற்றுமுறை அனைத்து வகுப்பறைகளிலும் கொண்டு வரப்பட வேண்டும். அன்றைய ஓலைச்சுவடி முதல் இன்றைய இணையம் வரை தகவல் பகிர்வின் பரிணாம வளர்ச்சியை எடுத்துக்காட்ட வேண்டும். தமிழில் கற்பிக்க மிக மிக எளிமையான சொற்களையே பயன்படுத்துதல் வேண்டும். கற்பவருக்குக் கடினமில்லாத,

பழக்கப்பட்ட சொற்களாக இருந்தால் அதுவே கற்ற கல்வியை தெளிவாய்ச் சிந்தை உள்வாங்கி உரிய பயனை அடையும்.

இலக்கியங்களை அசைவூட்டப்படும் வரைகலை (Animation) நுணுக்கங்கள் வாயிலாகப் பாடச்செய்திகளுடன் அளிக்கலாம். அகப்பொருள், புறப்பொருள் திணைகளை 99 பூக்கள் கொண்ட படங்களின் வாயிலாகக் கற்பித்து (கற்க நிற்க-வலைதளம்)மாணவரின் அறிவுத்திறனை வளர்க்கலாம்.



"பழையன கழிதலும் புதியன புகுதலும்" என்பதற்கேற்ப பத்துறையில் தோன்றி வரும் மாற்றங்களாலும் வளர்ச்சியாலும் தமிழ் போன்ற நெடிய பாரம்பரியத்தினைப் பெற்று விளங்கும் மொழியில் சூழ்நிலைகளால் ஏற்படும் மாற்றங்களை ஏற்றுக் கொண்டு தமிழ் வளர்ச்சிப் போக்கினை கணினியின் துணை கொண்டு மேம்படுத்த உலகளாவிய தமிழர்கள் அனைவரும் பாடுபடவேண்டும்.

மழலைக்கல்வி-தமிழ் இணையப் பல்கலைக்கழகம்வாயிலாகத் தமிழ் கற்பித்தல்

ஓர் ஆய்வு

புலவர். பி.ஆர். இலட்சுமி,

பி.லிட்.,எம்.ஏ.,எம்.ஏ., எம்ஃபில்.,

முனைவர் பட்ட ஆய்வாளர்

சென்னை-33.

"அருங்காட்சியகத்தில் இறந்த பொருட்களை வைத்திருப்பது போலக் குழந்தைகளை வகுப்பறையில் உட்கார வைத்துப் பூக்களின் மேல் புயல் வீசுவது போலப் பாடங்களைக் குழந்தைகளின் மேல் திணிப்பது தவிர்க்கப்படுதல் வேண்டும்."

- இரவீந்திரநாத் தாகூர்

தமிழ்மொழி கற்பித்தலில் உச்சரிக்கக் கற்பித்தலும், வரி வடிவத்தைப் பிழையற எழுதக் கற்பித்தலும் முக்கிய இடம் பெறுகின்றன.மேலும் தமிழ்மொழி கற்பித்தலில் கேட்டல்,பேசுதல்,படித்தல், எழுதுதல் திறன்கள் முக்கிய பங்கினை வகிக்கின்றன.இந்நூல்வகைத்திறன்களில் கேட்டல்,பேசுதல் திறனில் காது,வாய்,மூளை போன்ற உறுப்புகள் செயலாற்றுகின்றன. இத்திறன்களில் மாணவர்கள் எளிமையாகக் கற்கின்றனர்.படித்தல்,எழுதுதல் திறனில் கண்,கை,வாய்,மூளை இவை அனைத்தும் ஒருங்கிணைந்து செயல்படவேண்டியிருக்கும்.இதனால் எழுத்துகளின் வரிவடிவங்களைக் கற்க வேண்டிய அச்ச உணர்வு ஏற்பட்டுத் தமிழ் மொழியினை விலக்கும் நிலை உருவாகியுள்ளது. வகுப்பறையில் தமிழ் இணையப் பல்கலைக்கழகம் வாயிலாக மழலைக்கல்வியினை எளிதாகக் கற்பிக்கும் முறைதனை விளக்குவது இவ்வாய்வுக்கட்டுரையின் நோக்கம்.

புலம் பெயர்ந்தும்,பணியின் காரணமாகப் பன்னாடுகளில் வாழும் தமிழர் யாவரும் தமிழ் கற்க, பட்டம் பெற,தமிழ்ப்பண்பாட்டை அறிந்துகொள்ளத் துணைசெய்யும் நோக்கில் அமைக்கப்பட்ட தமிழ் இணையப் பல்கலைக்கழகம் ஏறத்தாழ 17 நாடுகளில் ஒருங்கிணைப்பு மையங்கள் நிறுவி பல்லாடகங்களின் துணை கொண்டு இணையவழித்தமிழ்க்கல்வியை வழங்குகிறது. கல்வித்துறையில் கணினி வழியில் மொழி கற்பித்தலைத் தகவல் தொடர்புத்தொழில்நுட்பத்துடன் இணைத்துத்தமிழுக்கெனத் தனி ஒரு உலகத்தை உருவாக்கியுள்ளது என்றால் அது மிகையாகாது.கற்பித்தலில் புதுமையைத் திட்டமிட்டுப் பாடப்பொருள் வழி கணினிக்கல்வி அளிப்பதன் வாயிலாக மாணவர்களின் கற்றல் நிலை மேம்பட வழியுண்டு. காலமாற்றத்திற்கேற்பவும்,வளர்ந்து வரும் நவீனத் தொழில்நுட்ப வசதிகளுக்கு ஏற்பவும் கல்வி முறையில் புதுமையாக்கத்தை மேற்கொண்டால் பண்பிலும்,அறிவிலும் சிறந்த வருங்காலத் தலைமுறையினரை உருவாக்க இயலும்.விதிமுறைகளை உருவாக்குவது முக்கியமல்ல.எப்படி அவ்விதி முறைகளைச் சிந்தித்துச் செயலாற்றுகிறோம் என்பதனை மனதில் கொண்டு கற்பித்தலின் சிறப்பு முறைகளான

- படி நிலைகளாகப்பிரித்தல்
- வடிவமைப்பு முறை
- தூண்டல் முறை

- குறிப்பு முறை
 - உடல்வழி
 - வாய்மொழி
- வரிசை முறைபோன்ற உத்திகளைப் பின்பற்ற வேண்டும்.

கருத்து வளர்ச்சியடையக் கற்பிக்கும் படிநிலைகளைத் தமிழ்இணையவழிக்கல்வி

- பொருத்துதல்
- அடையாளம் காட்டுதல்
- பெயர்கூறுதல்
- காட்சிப்படுத்துதல்

போன்றவற்றின் மூலம் தெளிவாக்கியுள்ளது.

பிற மொழிப்பாடங்கள் உலகில் பிழைக்கக் கற்றுக் கொடுக்கின்றன.தமிழ் மொழி மட்டுமே சிந்திக்கவும், வாழவும் கற்றுக் கொடுக்கிறது.கற்றலின் நோக்கங்கள் அறிவைப்பெறுவதற்காகக்கற்றல், செயல் புரியக்கற்றல், கூட்டாக வாழக்கற்றல், சுயம்இழக்காமல் இருக்கக் கற்றல் போன்றவையாகும்.இதற்கு வழிகாட்டியாகத் தமிழ் இணையக்கல்வி வலைதளமான (<http://www.tamilvu.org/courses/primer/bpooooo/.htm>.) இல் அமைந்துள்ள மழலைக்கல்வி முறையினைப் பயன்படுத்தலாம்.

"எண்ணும் எழுத்தும் கண்ணெனத் தகும்" என்ற சான்றோர் கருத்தின்படி தமிழ் இணையப்பல் கலைக்கழகம் சிறப்பாகச் செயலாற்றி வருகிறது.அமெரிக்கா உள்ளிட்ட பல்வேறு வெளிநாடுகளில் தமிழ்மொழி இணையப்பல்கலைக்கழகம் வாயிலாகக் கற்பிக்கப்படுகிறது. நால்வகைத் திறன்களை வளர்க்கும் விதமாக

1--	பாடல்கள்	--	12	பாடல்கள்(பயிற்சிகளுடன்)
2--	கதைகள்	---	4	கதைகள்(நீதிக்கருத்துகளுடன்)
3--	கதைகள்	--	4	கதைகள்(நீதிக்கருத்து மற்றும் பயிற்சிகளுடன்)
4---	உரையாடல்	--	7	தலைப்புகள்(பயனுள்ள நடைமுறைக்கருத்துகள்)
5---	வழக்குச்சொற்கள்	--	10	பாடல்கள்(வாழும் உலகியற் பொருட்கள்,மரபுச்சொற்கள்)
6----	நிகழ்ச்சிகள்	---	9	பாடங்கள்(காலங்கள்)
7----	எண்கள்	----	9	பாடங்கள்(பயிற்சி,பாடம்,பாடல்கள்)
8----	எழுத்து	----	14	பாடங்கள்(ஒரெழுத்து முதல் ஐந்தெழுத்துச்சொல் வரை அறிதல், பாடல்கள்)

பாடம், பயிற்சி என வகுக்கப்பட்டு அசைவூட்டப்பட்ட வரைபடங்களுடன் தெளிவாகவும், சிறந்த ஒளியமைப்புடனும் அமைக்கப்பட்டுள்ளது. இதன் மூலம் கற்றல்திறன் சிறப்பாக அமையும்.

பாடல்கள் வரிசையில் 'காக்கைப்பாடல்','ஆத்திசூடி' இரண்டு முறை இடம்பெற்றுள்ளன.

தமிழுக்கு உரிய சிறப்பு ஒலி எழுத்துகள்-ல,ள,ழ,ன,ந,ண,ர,ற
 ஒரே மாதிரித்தோற்றமளிக்கும் எழுத்துகள்-அ,ஆ,ஒ,ஓ,ஔ
 ஒன்று போல அமையும் எழுத்துகள்-எ,ஏ மற்றும் உ,ஊ
 இரு வரிவடிவ எழுத்துகள்-கெ,கே,கை
 மூன்று வடிவ எழுத்துகள்-கொ,கோ, கௌ
 துணை எழுத்துகள் இணைத்து எழுதும் எழுத்துகள்-கா,கை

கீழ் வளைவு-பு,பூ

மேல் வளைவு-மி,மீ

கீழ்விலங்கு-கு,கூ

மேல்விலங்கு--கெ,கொ,கே,கோ

அழுத்தம் கொடுத்து உச்சரிக்க வேண்டியவை-ச,ண,ற,ள

எளிய முறையில் உச்சரிக்க வேண்டியவை--ன,ர,ல

நாவை வளைத்து உச்சரிக்க வேண்டியவை-ழ

பழக்கத்தில் கற்றுக் கொள்ள வேண்டியவை-ங,ஞ

மேற்கண்ட செய்திகள் அனைத்தையும் இணைத்து மழலைக்கல்வியில் இணைத்துக் கற்பிக்கலாம்.

'பள்ளி' என்ற தலைப்பில் படத்தைச் சுட்டிப் பொருத்தமான எழுத்தை நிறைவு செய்யும் பகுதியும், எழுத்தைக்கூட்டி ஒலிக்கும் முறையும் (<http://www.tamilunltd.com/swfs/mainpage.html>) வலைதளத்தில் சிறப்பாக அமைக்கப்பட்டுள்ளது. (<http://www.kidsone.in/tamil/>) என்ற வலைதளத்தில் மெய்யெழுத்துகள் எழுதும் முறைகளும்,பயிற்சித்தாள் அமைக்கப்பட்டுள்ள விதமும் சிறப்பானது.இவ்வாறு பிற வலைதளங்களில் உள்ள சிறப்பான பாடங்களைப் பயன்படுத்தலாம். (www.tamilvu.org/courses/beginner/main.htm.) என்ற இணையதளத்தில்

- மெய்யெழுத்து அறிமுகப் பாடல்(அக்காள் வீட்டுத்தோட்டம்)
- இன எழுத்துப்பாடல்(எங்கள் அம்மா இனியவள்)
- எனகர,ணகர வேறுபாடு அறியும் பாடல்(மன்னன் நல்ல மன்னன்)
- உயிரெழுத்து அறிமுகப்பாடல்(அன்பே கடவுள் அறிவாய் பாப்பா) முதலியவை சிறப்பானவை.

(www.tamilvu.org/courses/beginner/main.htm.)என்ற இணையதளத்தில் அமைந்துள்ள எழுத்துகளை எழுதும் பயிற்சிகள் போன்றவற்றை மழலைக்கல்வியில் இணைத்துக் கற்றுக்கொடுக்கலாம்.

(<http://www.tamilvu.org/courses/primer/bpooooo/.htm>.) என்ற வலைதளத்தில் 'அணிலும் ஆடும்' பாடலில் 'ணி' என்ற எழுத்து தவறாகக் குறிக்கப்பட்டுள்ளது. அசைவூட்டப்பட்ட வரைகலைப்படங்களை இன்னமும் தெளிவாக அமைத்தால் சிறப்பாக இருக்கும்.

பழங்காலப் பயிற்சிமுறை இன்றைய மாணவர்களுக்கு நடைமுறைப் பயன்பாட்டில் சிக்கல்களை ஏற்படுத்தும் பொழுது புதிய கருத்தாக்கங்கள் தோன்றுவது இயல்பு.இவ்வாறு தமிழ்இணையப் பல்கலைக்கழகம் பாடத்திட்டம், பயிற்றுமுறை,மதிப்பீடு என யாவற்றிலும் தமிழ் கற்பித்தலைச் சிறப்பாக்கியுள்ளது."உன் ஆசிரியர் என்ன படித்திருக்கிறார்?எனக் கேட்பதை விட உன் ஆசிரியர் என்ன படித்துக்கொண்டிருக்கிறார்?" என்ற கல்வியாளரின் சிந்தனை ஒவ்வொரு ஆசிரியரிடமும் இன்றைய நாளில் எதிர்பார்க்கக்கூடிய ஒன்று. இதன்மூலம் நடைமுறை மாற்றப் புதிய கல்விக்கோட்பாடுகளைப் புகுத்தமுடியும்.

வகுப்பறைச் சூழலில் தமிழ் மொழியைக் கற்பித்தலில்

கணினித் தொழில்நுட்பத்தின் பயன்பாடு : ஓர் கள ஆய்வு

அ.கோவிந்தராஜ்

ஆங்கில மொழி விரிவுரையாளர்,
இந்திய நிர்வாகவியல் துறை,
மீனாட்சி பல்கலைக்கழகம், சென்னை -78.
Email: govindias@gmail.com

‘பழையன கழிதலும் புதியன புகுதலும், வழுவல கால வகையினான” என்பார் நன்னூலார். புதுமை எண்ணமும், அறிவியல் வளர்ச்சியும் காலந்தோறும் கற்றல் கற்பித்தல் முறைகளில் பல்வேறு மாற்றங்களை ஏற்படுத்தி வருகின்றன. குருகுல முறை கற்பித்தலில் வாய்மொழி வழக்காய் இருந்து பின்னர் கரும்பலகை வழக்காய் மாறி, இன்று கணினியின் துணை கொண்டு கற்பித்தல் என்ற ஒரு நிலை உருவாகியுள்ளது. காலத்தின் கோலத்திற்கேற்பப் பல்வேறு பாடப்பிரிவினைக் கற்பிக்கின்ற ஆசிரியர்கள் தங்களின் கற்பிக்கும் முறையினை மாற்றிக் கொண்டே வந்துள்ளனர். ஆனால், வகுப்பறைச் சூழலில் தமிழ் மொழியைக் கற்பித்தலைப் பொருத்தமட்டில் கணினியின் துணை கொண்டு கற்பித்தல் என்பது அருகியே காணப்படுகின்றது. தமிழ் மொழியைக் கணினித் தொழில்நுட்பம் கொண்டு கற்பிப்பதில் காணலாகும் இத்தகு தேக்க நிலைக்கான காரணத்தையும், தேக்கநிலையை நீக்குவதற்குரிய வழிவகைகளையும், தேக்கநிலை நீங்கினால் ஏற்படும் பயன்களையும் இக்கட்டுரை முன்வைக்கின்றது.

கள ஆய்வு

தமிழ் மொழியைக் கற்பிக்க வகுப்பறையில் கணினித் தொழில்நுட்பத்தைப் பயன்படுத்துவதற்கான சாத்தியக் கூறுகள் உள்ளனவா? அவ்வாறு பயன்படுத்த முனையும் போது எதிர் கொள்ளும் நடைமுறைச் சிக்கல்கள் யாவை? கணினித் தொழில்நுட்பத்தைப் பயன்படுத்திக் கற்பிக்கும் பொழுது மாணவர்களின் தமிழ் மொழிப் பயன்பாடு மேன்மையடையுமா? என்பன போன்ற வினாக்களுக்கான விடையை அறியவும் அதன் நம்பகத்தன்மையை உறுதி செய்யவும் ஓர் கள ஆய்வு மேற்கொள்ளப்பட்டுள்ளது.

ஆய்வுக்களம்

சென்னை மாநகராட்சிப் பகுதிக்கு உட்பட்ட 10 பள்ளிகளில் இக்கள ஆய்வு நடத்தப்பட்டது. 10 பள்ளிகளில் ஐந்து அரசு சார் பள்ளிகள், ஐந்து தனியார் பள்ளிகள். இப் பள்ளிகளில் பணிபுரிகின்ற பத்தாம் வகுப்பிற்குப் பொதுத் தமிழ் பாடம் கற்பிக்கின்ற 25 தமிழ் ஆசிரிய ஆசிரியைகள் இவ்வாய்வில் பங்கேற்றுள்ளனர்.

ஆய்வு அணுகுமுறை

ஆய்வின் போது தேவையான தரவுகளைத் (data) திரட்டுவதற்கு இரண்டு ஆய்வு அணுகு முறைகள் பின்பற்றப்பட்டுள்ளன. முதலாமாவது நேர்காணல் (Interview) அணுகுமுறையாகும். இவ்வணுகு முறைக்கு வலுச்சேர்க்கின்ற வகையில் இரண்டாவதாக வினாத்தாள் (Questionnaire) அணுகுமுறையும் இவ்வாய்வில் பயன்படுத்தப்பட்டுள்ளது.

ஆய்வில் கண்டறியப்பட்ட உண்மைகள்

கள ஆய்வின் போது திரட்டப்பட்ட தரவுகளின் (Data) அடிப்படையில் கண்டறியப்பட்ட உண்மைகள் வருமாறு:

- ஆய்வு மேற்கொள்ளப்பட்ட அனைத்து பள்ளிகளிலும் ஆசிரியர்களின் பயன்பாட்டிற்காக குறைந்தது இரண்டு கணினிகள் ஒதுக்கப்பட்டுள்ளன.
- ஆய்விற்கு உட்படுத்தப்பட்ட அனைத்து ஆசிரிய ஆசிரியைகளும் (100 சதவிகிதம்) தமிழ் மொழியைக் கற்பிக்கக் கணினித் தொழில்நுட்பத்தைப் பயன்படுத்த முடியும் என்று கூறியுள்ளனர்.
- ஆனால் ஆய்விற்கு உட்படுத்தப்பட்டவர்களில் 80 சதவிகிதம் பேர் மட்டுமே கணினியைக் கையாளத் தெரியும் என்று கூறியுள்ளனர்.
- மேலும் ஆய்விற்கு உட்படுத்தப்பட்டவர்களில் 20 சதவிகிதம் பேர் மட்டுமே தமிழ் மொழியைக் கற்பிக்க கணினித் தொழில்நுட்பத்தைப் பயன்படுத்துகின்றனர்.
- ஆய்விற்கு உட்படுத்தப்பட்ட அனைவரும் கணினித் தொழில்நுட்பத்தின் துணை கொண்டு கற்பிப்பதால் மாணவர்களின் தமிழ் மொழிப் பயன்பாடு மேன்மையடையும் என்று கூறியுள்ளனர்.
- ஆய்வு மேற்கொள்ளப்பட்ட பள்ளிகளில் 70 சதவிகிதம் பள்ளிகளில் இணைய தள வசதியோ (Internet - இன்டர்நெட்) பல்லாடக ஆய்வக (Multimedia lab - மல்டிமீடியா) வசதியோ இல்லை.
- ஆய்விற்கு உட்படுத்தப்பட்டவர்கள் அனைவரும் இணைய தளம் மற்றும் பல்லாடக ஆய்வகம் போன்றவற்றின் துணை கொண்டு தமிழ் மொழியைக் கற்பிக்க முடியும் என்று கூறியுள்ளனர். ஆனால் இவர்களில் 8 சதவிகிதம் பேர் மட்டுமே இதுவரை இணைய தளத்தினைப் பயன்படுத்தியுள்ளனர்.
- பல்லாடகம் மற்றும் இணையதளத்தின் உதவிகொண்டு தமிழ் மொழியைக் கற்பிப்பதால் மாணவர்களின் கருத்தை எளிதில் கவர முடியும் என்று அனைத்துப் பங்கேற்பாளர்களும் கூறியுள்ளனர்.

கணினித் தொழில்நுட்பத்தை வகுப்பறைச் சூழலில் பயன்படுத்த முடியாமைக்கான காரணங்களாகப் பங்கேற்பாளர்கள் முன்வைத்தவை:

1. போதிய தொழில்நுட்ப அறிவின்மை: ஆய்வில் பங்கேற்றவர்களில் அனைவரும் தமிழ் மொழியைக் கற்பிக்கக் கணினித் தொழில்நுட்பத்தைப் பயன்படுத்த வேண்டும் என்ற அவாவை வெளிப்படுத்தியிருந்தாலும் அவர்களுக்குக் கணினி தொடர்பான போதிய தொழில் நுட்ப அறிவின்மையை மிகப் பெரிய தடைக்கல்லாக முன்வைத்தனர்.
2. போதிய அளவு மென்பொருட்கள் (Software) இன்மை: ஆய்வில் பங்கேற்றவர்களில் ஒருசிலர் கணினித் தொழில்நுட்பத்தினைக் கற்றிருந்தாலும் அவர்கள் பயன்படுத்தும் வகையில் தமிழ் மொழி கற்பிக்கும் மென்பொருட்கள் இல்லாமல் இருப்பதை ஒரு தடைக்கல்லாகவே கருதுகின்றனர்.
3. போதிய உட்கட்டமைப்பு (Infrastructure) வசதியின்மை: வகுப்பறைகள் மற்றும் மாணவர்களின் எண்ணிக்கையை ஒப்பிடும் போது போதிய அளவு கணினி வசதியோ, பல்லாடக ஆய்வக (multimedia lab) வசதியோ இல்லாமல் இருப்பதை ஆய்வில் பங்கேற்றவர்கள் ஓர் தடைக்கல்லாகக் கருதுகின்றனர்.
4. நேரமின்மை : பெரும்பாலான பள்ளிகளில் தமிழ்மொழிக்கு முக்கியத்துவம் கொடுப்பது அரிதாகவே உள்ளது. தமிழ் வகுப்பிற்கானப் பாட வேளைகளைப் பெரும்பாலும் குறைத்து விடுகின்றனர். ஒரு சில பள்ளிகளில் தமிழ் வகுப்பினைப் பிற பாட ஆசிரியர்களுக்குக் கொடுக்கும்படி வற்புறுத்தப்படுகின்றனர். இதனால் கிடைக்கும் குறுகிய கால அளவில் தமிழாசிரியர்கள் பெரும்பாலும் பாடத்திட்டத்தினை (syllabus) முடிப்பதிலேயே குறியாக இருக்க வேண்டியிருப்பதாகவும், கணினிசார் தொழில் நுட்பத்தைப் பற்றி சிந்திக்கவே நேரமில்லை என்பதையும் முன் வைத்துள்ளனர்.

கணினித் தொழில்நுட்பத்தின் துணை கொண்டு கற்பிப்பதால் விளையும் பயன்களாகப் பங்கேற்பாளர்கள் பட்டியலிட்டவற்றுள் சில:

- கண்ணால் காண்பதும், காதல் கேட்பதும் ஒரே சமயத்தில் நிகழ்வதால் மாணவர்களின் கவனச் சிதைவு குறையும், மாணவர்களுக்குப் பாடம் எளிதில் புரியும்.
- ‘கற்றது உணர விரித்துரையாதார் - வாசயில்லா மலராவார்” (குறள்: 650) என்பார் வள்ளுவர். கணினித் தொழில்நுட்பத்தின் துணை கொண்டு கற்பிப்பதால் ஆசிரியர் தான் உணர்ந்தவற்றைத் தெளிவாக எடுத்துரைக்கும் வாய்ப்புக்கிட்டும்.
- கற்றல் சூழலில் இறுக்கநிலை ஏற்படும் போது, கணினிசார் தொழில்நுட்ப வழி கற்பித்தல் ஓர் நெகிழ்ச்சி நிலையை ஏற்படுத்திச் சூழல் மாற்றம் காரணமாக மாணவர்களின் தொடர் கவனத்தை அதிகரிக்கும் இதனால் அவர்களின் கொள்திறனும், ஆக்கத் திறனும் பெருகும்.
- பாடப் பொருளில் புதிய பார்வையை உருவாக்கும்.
- ஆசிரியருக்கான பணிச்சுமை குறையும். அதன் விளைவாக அதிக நேரத்தினை மாணவர்களின் ஐயங்களை நீக்கப் பயன்படுத்தலாம்.

பங்கேற்பாளர்களின் ஐயங்களும், அவற்றுக்கான விளக்கங்களும்:

கள ஆய்வின் போது பங்கேற்பாளர்கள் சில ஐயங்களை ஆய்வாளரிடம் எழுப்பினர். அவ்வையங்களை நீக்கும் பொருட்டு ஆய்வாளர் அளித்த விளக்கம் பின்வருமாறு:

ஐயம் 1 : தமிழ் மொழியைக் கற்பிக்க கணினியின் துணை தேவைதானா?

விளக்கம்: “கரும்பலகை எழுத்தும், வாய்மொழி விளக்கமும்” ((chalk and talk method)) என்ற அணுகுமுறையே பல்லாண்டு காலமாகத் தமிழ்மொழியைக் கற்பிக்கப்பயன்படுத்தப்பட்டு வருகின்றது. இன்றைய சூழலில் இவ்வணுகுமுறையானது தமிழ்மொழி மீது மாணவர்களுக்கு ஓர் ஈர்ப்பினையோ, பிடிப்பினையோ ஏற்படுத்தத் தவறுகிறது. கணினி மீதும் மற்றும் அது சார்ந்த தொழில்நுட்பத்தின் மீதும் இன்றைய தலைமுறையினருக்கு இருக்கும் ஓர் ஈர்ப்பு சக்தியே அவற்றை மாற்றும் சக்தியாக உள்ளது. இந்த ஈர்ப்பு சக்தியை தமிழ் ஆசிரியர்கள் முறையாகப் பயன்படுத்தினால், மாணவர்களுக்குத் தமிழ்மொழி மீது ஓர் தீராக் காதலை ஏற்படுத்தலாம். அதன் பொருட்டே இன்றைய சூழலில் தமிழ் மொழியைக் கற்பிக்க கணினியின் துணை தேவையகிறது.

ஐயம் 2 : அனைத்துப்பாடங்களுக்கும் அனைத்து வகுப்புகளிலும் கணினித் தொழில் நுட்பத்தைப் பயன்படுத்துவது சாத்தியமா?

விளக்கம் : இன்றைய நிலையில் கணினியின் விலையும் அது சார்ந்த தொழில்நுட்பங்களின் விலையும் அதிகமாக இருப்பதே மேற்கண்ட ஐயப்பாட்டிற்குக் காரணமாக இருக்கும் என்று தோன்றுகிறது. அப்படியிருக்குமாயின், காலத்தின் ஓட்டத்தில் இந்நிலை மாறும். 1970 களில் அருகிக் காணப்பட்ட தொலைக்காட்சிப் பெட்டி, 1980களில் ஊருக்கு ஒன்று என்றும், 1990களில் வீதிக்கு ஒன்று என்றும், 2000 களில் வீட்டிற்கு வீடு என்றும், 2005க்குப் பின்னர் அறைக்கு ஒன்று என்றும், எங்கும் நீக்கமற நிறைந்துள்ளது. இதே போன்று இன்று பள்ளிக்கு ஒரு கணினி என்பது நாளை வகுப்பிற்கு ஒரு கணினி என்றும், நாளாவட்டத்தில் மாணவனுக்கு ஒரு கணினி என்றும் காலம் மாறும், எனவே தமிழ் மொழியைத் கற்பிக்க கணினித் தொழில்நுட்பத்தினைப் பயன்படுத்துவது சாத்தியமே.

ஐயம் 3 :

ஒவ்வொரு ஆசிரியரும் தனித்தனியே பாடத்திட்டங்களை உருவாக்க முயலும் போது அவர்களுக்குக்கால விரயமும், பண விரயமும் ஏற்படுமே?

விளக்கம் :

'Core Banking' என்ற தொழில்நுட்ப வளர்ச்சி இன்று உலகில உள்ள வங்கிகள் அனைத்தையும் ஒன்றிணைத்து வங்கி நடவடிக்கைகளை எளிமையானதாகவும் வேகமானதாகவும் மாற்றியுள்ளது. அதே போல் 'Core Teaching' என்ற ஒரு தொழில்நுட்பப் பிரிவினை ஏற்படுத்தி தமிழ்நாட்டில் உள்ள அனைத்துப் பள்ளிகளையும் இணைக்கலாம். தமிழ்நாட்டுப் பாட நூல் நிறுவனம் ஒவ்வொரு வகுப்பிற்கான பாடத்திட்டங்களை உருவாக்கும் போதே கணினியின் வாயிலாக கற்பிக்க ஏதுவான பாடங்களைக் கண்டறிந்து, அதற்கு தேவையானவற்றை உருவாக்க ஆசிரியர்கள் மற்றும் கணினித் தொழில்நுட்ப வல்லுனர்கள் கொண்ட குழு ஒன்றினை அமைக்க வேண்டும். இவர்கள் கூட்டு முயற்சியில் உருவாகும் மென்பொருட்களை ஆசிரியர்கள் பதிவிறக்கம் (எழறடெழயன்) செய்து பயன்படுத்த முற்பட்டால், பண விரயமும், கால விரயமும் தவிர்க்கப்படும்.

ஐயம் 4: கணினித் தொழில்நுட்பமானது வகுப்பறையில் பாடம் நடத்தப் புகுந்துவிட்டால், ஆசிரியரின் நிலை என்ன? ஆசிரியரைக் கணினி ஈடு செய்யுமா?

விளக்கம் : ஆசிரியன் ஓர் உயிரோட்டமுள்ள ஓவியன். இவ்வோவியன் உருவாக்கும் ஓவியமே மாணவன். ஓவியனின் வெளிப்பாட்டுத்திறனை அதிகரிக்கவும், ஓவியத்தின் அழகை மெருகூட்டவும் உதவும் தூரிகையே கணினி. கணினி பாடம் நடத்த உறுதுணைபுரியும் ஒரு கருவியே (download) அன்றி அது ஒரு ஓவியன் அன்று.

செயல் விளக்கம் :

ஆய்வாளர் கள ஆய்வோடு நில்லாமல் கணினி சார்ந்த தொழில்நுட்பத்தினைப் பயன்படுத்தித் தமிழ் மொழியைக் கற்பிக்க முடியும் என்பதனை ஒரு சங்கப்பாடலைத் தெரிவு செய்து நிலைநிறுத்தியுள்ளார். தமிழ்நாடு அரசின் பத்தாம் வகுப்புப் பாடத்திட்டத்தில் உள்ள கபிலர் இயற்றிய 'வேரல் வேலி...' எனத் தொடங்கும் குறுந்தொகைப் பாடலை (எண்:18) எவ்வாறு கணினித் தொழில்நுட்பத்தின் மூலம் மாணவர்களின் கருத்தைக் கவரும் வண்ணம் வகுப்பறையில் கற்பிக்கலாம் என்பதனைக் கள ஆய்வின் போது ஆய்வாளர் ஆசிரியர்களிடம் செயல் முறையில் விளக்கியுள்ளார். Pழறநச பழிவெ பசநளநவெயவழைந என்று சொல்லக் கூடிய கணினித் தொழில்நுட்பத்தைக் கொண்டு குறிஞ்சித் திணையைக் காட்சிப்படுத்தியும், Coral draw, Photoshop மற்றும் Animation போன்ற தொழில்நுட்பத்தைப் பயன்படுத்தி வேர்க்கோட்டுப் பலா மரத்தினைச் சுற்றி மூங்கில் (வேரல்) வேலி அமைந்திருப்பதையும், இவ்வியற்கைச் சூழல், தலைவிக்கு தலைவன் வேலியாய் அமைய வேண்டும் என்பதை நினைவூட்டுவதையும் மாணவர்களுக்கு எளிதில் புரியும்படி ஆய்வாளர் காட்சிப்படுத்தியுள்ளார். மேலும் குடியாளர் என்னும் தொழில்நுட்பத்தைக் கொண்டு அப்பாடலில் அமைந்துள்ள மோனையை (வேரல் : சாரல்) மாணவர்களின் மனதில் நிற்கும் வண்ணம் ஒலி மற்றும் ஒளிப்படுத்தியுள்ளார். மேலும் செய்யுளுக்குரிய ஓசையோடு அறியச் செய்தல், செய்யுளின் பொருள் உணர்ச்சிக்கு ஏற்ற குரல் ஏற்றத் தாழ்வுடன் படிக்கச் செய்தல் போன்ற கூறுகளுக்கும் செயல் முறை விளக்கத்தில் ஆய்வாளர் முக்கியத்துவம் கொடுத்திருந்தார். இச்செயல் விளக்கத்தினைக் கண்ட அனைத்து பங்கேற்பாளர்களும், தாங்களும் இது போன்றதொரு புதிய அனுபவத்தை மாணவர்களுக்கு அளிக்க முற்படுவோம் என்று ஆய்வாளரிடம் தெரிவித்துள்ளனர்.

ஒவ்வொரு கால கட்டத்திலும் ஒவ்வொரு ஊடகத்தின் வழித் தமிழ் தன் பயணத்தை இனிதே மேற்கொண்டுள்ளது. தர்க்கத்தின் தோளேறியும், சமயத்தின் தோளேறியும், காலத்தைப் புறத்தள்ளி வந்துள்ளது தமிழ். அவ்வகையில் நடப்பு நூற்றாண்டோ 'நாளும் இன்னிசையால் நல்ல தமிழ் வளர்ப்போம்' என்ற நிலையோடு 'கணினித் தொழில் நுட்பத்தாலும் இனிதே தமிழ் வளர்ப்போம்' என்ற நிலை உருவாகியுள்ளது. இந்நிலையில் வகுப்பறைச் சூழலில் தமிழ் மொழியைக் கற்பிக்கக் கணினித் தொழில் நுட்பத்தின் துணையை நாடத் தமிழ் ஆசிரியர்கள் முற்பட வேண்டும். ஆசிரியர் பயிற்சி படிப்பின் (Teacher

Training) போதும், இளங்கலை, மற்றும் முதுகலைத் தமிழ் இலக்கியப் படிப்பின் போதும் அடிப்படைக் கணினி அறிவு மாணவர்களுக்கு வழங்கப்பட வேண்டும். இம் மாணவர்களே நாளைய ஆசிரியர்கள். இவர்களுக்கு அடிப்படைக் கணினிசார் தொழில்நுட்ப அறிவு மற்றும் பயிற்சி வழங்கப்படுமாயின், எதிர்காலத்தில் ஆசிரியர்களாக வரும் இவர்கள், வகுப்பறைகளில் கணினி சார் தொழில்நுட்பத்தினைப் பயன்படுத்திட முனைவர். ஏற்கனவே பணியில் உள்ள ஆசிரியர்களுக்குப் புத்தாக்கப்பயிற்சி (Refresher Course) யின் போது அடிப்படைக் கணினி சார் தொழில் நுட்பப் பயிற்சி விரிவாக அளிக்கப்பட வேண்டும். மேலும் தமிழக அரசானது 'உத்தமம்' (உலகத்தமிழ் தகவல் தொழில் நுட்ப மன்றம்) போன்ற அரசு சாரா அமைப்புகளின் துணை கொண்டு தமிழ் மொழியைக் கற்பிக்க உதவும் பல புதிய மென் பொருட்களை உருவாக்க முனைய வேண்டும். மேற்கண்ட எண்ணங்கள் யாவும் செயல் வடிவம் பெறுமாயின்,தமிழ் மொழியின் வளர்ச்சிக்குக் கணினித் தொழில் நுட்பத்தின் துணையை நாடும் இன்றைய நிலை மாறி, கணினித் தொழில் நுட்பத்தின் வளர்ச்சிக்குத் தமிழ் மொழியின் துணையை நாடும் நிலை நாளை தோன்றும் என்பது திண்ணம்.

உறுதுணை நூல்கள்

1. முப்பாலிகை, அ. 'தமிழின் எதிர்காலம்'.
2. கருணாகரன், கி. 'மொழி வளர்ச்சி'.
3. வரதராசன், மு. 'மொழி வரலாறு'.
4. Govindaraj, A. 'Importance of Inch - Box in English Language Teaching'. (Research Paper)

Open Educational Resources in the Context of Teaching and Learning of Tamil as the First Language

Dr.N.Balasubramanian

Director, School of Distance Education,
Bharathiar University
Coimbatore.
nbalu1954@yahoo.com

Introduction

There has been a paradigm shift in the role of Distance Education universally. Traditionally, distance education was restricted in terms of enrolment because of production, reproduction, and distribution costs. Even though it costs the university in terms of money and time to produce a course, technology has made the reproduction cost almost non-exist which supports the universities in the fulfillment of the promise of right to universal education. At no cost universities could make the content available to millions substantially improving the quality of the life of the learners around the world. The Government of India created its National Knowledge Commission as a high level advisory body to the Prime Minister with an objective of transforming India into a knowledge society. Conceived as a network, the Virtual University for small states of the Commonwealth paved the way for sharing materials and programmes due to the emerging trend of developing OERs which ensures education for all. At this juncture, the present paper aims at discussing how the OERs and other computer and web based technologies could be exploited in teaching a language particularly Tamil as the first language in general and particularly to NRIs of Tamil Nadu origin settled elsewhere in the world through Open and Distance Education system.

Open Educational Researches: An Asset for Open and Distance Learning

Open Educational Resources is a relatively new concept which may be seen as a trend towards openness in higher education including movements such as Open Source Software and Open Access. The two important aspects of openness involve free availability in the internet and little restrictions on the use of the resources. The end-user should be free from technical barriers, price barriers and legal barriers on the part of the resources. Besides using the materials, the end user is totally free to adopt, build upon and reuse the same provided the original creator is acknowledged for the work. The participants of a UNESCO Forum in 2002 defined OER as “the open provision of educational resources, enabled by information and communication technologies for consultation, use and adoption by a community of users for non-commercial purposes”. It is also defined as “OERs are digitized materials offered freely and openly for educators and self-learners. The use and reuse for teaching, learning and research”. The OERS may include (i) Learning Content: full courses, coursework, content modules, learning objects, collections and journals (ii) Tools: software to support the development, use, reuse, and delivery of learning content and (iii) Implementation Resources: intellectual property licenses to promote design and localization of the learning materials.

Mapping OERS

It is to understand that a number of projects and initiatives in the field of OERs is growing fast which include institution based or institution supported initiatives. So far as OER Movements in post-secondary education is concerned, it is found that over 150 universities in China have produced over 450 courses online. The Paris Tech OCW Project formed by universities in France offers 150 courses. The Japanese OCW Alliance formed by nine most prestigious universities in Japan offers over 250 courses in Japanese language and 100 courses in English language. MIT, RICE, Johns Hopkins, Tufts, Carnegie and Utah State Universities in USA have large scale OER programmes. It has been estimated that altogether there are over 2000 university courses freely available online. Australia, Brazil, Canada, Hungary, India, Iran, Ireland, the Netherlands, Portugal, Russia, South Africa, Spain, Thailand, the UK, the USA and Vietnam have already started functioning in the field of OER.

Users and Producers of OER

The typical OER users are to be a single enthusiastic, a well educated self-learner, a faculty member, etc. The motive in involving OER movements may be that if Universities do not support the open sharing of research results and educational materials, traditional academic values may be marginalized by the market forces. It is also true that free sharing leads to broader and faster dissemination of knowledge which results in development of problem solving skill among the learners besides reducing social inequality. It also increases the popularity, reputation and the pleasure of sharing the resources with fellow users.

Challenges to the Growing OER Movement

Lack of awareness of copyright issues is commonly seen among the users of OERs. Open licensing may ensure controlled sharing with some rights reserved to the author. However, there has been a growing interest for open licenses which is witnessed by the increasing number of OERs under Creative Commons License. The users of OERs find problems in judging the quality and relevance of such materials. This problem could be managed if the producers use their brand or reputation of their institutions. A review report of the materials prepared by their peers will also help ensure quality. Sustainability of the OER initiatives is also a challenge in the field. Hence, it is important to seriously consider how such initiatives could be sustained in the long run.

Natural Language Processing and Language Learning and Teaching

In learning their mother tongue, children develop some remarkable sets of capabilities using which they acquire knowledge and skills enabling them to produce and comprehend an indefinite number of new utterances and judge their properties. Hence, we have to devise an explanatory account of the mental operations important for the development of linguistic abilities. At this juncture, it is imperative to seek technological applications to enhance such linguistic skills among learners.

One such technological application is the Natural Language Processing (NLP) supported by computing technology. NLP is concerned with the computational modeling and design and development of a wide variety of systems leading to non-machine communication. The NLP is nothing but the ability on the part of the computer to analyze written text or spoken utterances without limiting to mere recognition of graphics of phonemes but extending even up to the development of an understanding with reference to the task at hand besides integrating various modules with interactive layers. Natural language interface

to databases, computers question answering systems, story understanding, machine translations, etc., are some of the applications of NLP which could very well be exploited in teaching a language to children, particularly Tamil as the first language at the international level.

Software for Teaching of Reading & Writing

Lass (1981) identified a number of characteristics that typify good teachers of reading such as ability to organize and manage instruction, attend to individual needs, pace instruction correctly monitoring student attention, monitoring of their achievement, providing one to one instruction when needed and use supplementary materials. Software specially designed for enhancing readiness skills in reading, teaching decoding such as right vocabulary, phonics and structural analysis which include teaching of homonyms and synonyms, teaching of reading comprehension and study skills should be made popular among language teachers for effective accomplishment of language skills among children. Children could be introduced to word processing in order to improve their writing skill. Because of its advantages such as ease of correction, ease of revision, formatting ease, time & effort, quantity and quality of writing, etc. word processing could be the best tool in the hands of the teacher in developing the writing skill among children.

Multimedia and Author ware: Tools for Development of CALL Packages

Multimedia, being a potential tool in the hands of a language teacher, he could develop multimedia packages in different subjects. Even a teacher with little knowledge in computing could do so by using authoring tools such as Author ware, Hyper Cord, Hyper Studio, Mediator, Winwida, Question Mark, etc. Once becoming familiar with the way the authoring tool works, it is then a matter of design rather than programming.

Generally, there are two types of multimedia packages in language teaching, viz., *References* and *CALL Packages*. The *References* may include encyclopedias and dictionaries. Encyclopedias are the vast collection of hyper-texts, scanned photographs, animations, graphics, human voice, music and video clips. By providing menus and buttons, we can activate the programme and link various parts of it all together. *CALL* packages help the learner practice the language by providing models of the language in context and encouraging the user to imitate the example in a structured way. Using the record option, an inbuilt facility, the user may record his / her voice and compare it to that of the model and thus the course develop his pronounceate, stress, pausing, phrasing and intonatic.

E-Mail as a Tool for Teaching a Language

E-mail provides a medium for collaborative work using information collected from various locations, besides providing real purposes and audience in relation to the development of literary skills. E-mail helps the teacher introducing, managing and writing up projects due to the facility for sorting all communications. E-mail provides more authentic purposeful exposure to the language in hand and mire meaningful interactive when compared to the conventional classroom activities. This ultimately results in higher level student motivation because of the opportunity given to the learner to practice with the target language in open-ended situations. The writing skill also be improved because of the real purpose with the real people that centres round the activity. Being learner centred and individualized in learning, learners feel free in focusing their areas of interest. Multinational cross cultural discussions, intercultural authropology, review of text, film, etc may be better e-mail projects for the learners. The phenomenal

growth in e-mail use lead to greater internationalization and an increased demand for a world language. It is hoped that by adopting e-mail projects in the context of teaching and learning of Tamil, we can effectively teach the language skills among the learners.

Computerized Question Banks and Online Testing

The computer could be used as an aid to construct as well as administer tests in the context of teaching and learning of any language which enables utilizing the capabilities of the computer system to generate print and administer tests to the learners and score their performances. Networks of computer terminals ensure equally administration of the test directly right to the terminal. Once decoded, the items at the given specifications viz., type of questions, level of cognition, difficulty level and discriminative power, etc., could be printed, duplicated and distributed to the candidates who record their responses on machine-readable forms for scoring. It is possible to call the test items either in the same order as stored or even drawn at random before printing and duplicating. It is also desirable to call the items randomly for each student but printed in different order allowing the students receive a different set of questions. Implementation process of the testing system involves three natural phases, viz., before the test, during the test and after the test with defined roles both to the teacher and the students in each phase. It is to true that computer technology has come to the rescue of the teacher as well as the students from inhospitable regions of examinations since during testing situation both of them seen anxious. Online testing could also be exploited in the context of teaching and learning of a language since such tests are specially designed for the open and distance learners.

Conclusion

Education is the sector where the international countries have been pumping in enormous funds but still it has not reached the grass roots level of the community. However, it is obvious that everyone in the modern world has a fire to learn and educate oneself which is a healthy indication but often gets drained due to inadequate facilities. This block could be overcome with the emergence of Distance Education which opened a new window of opportunities to those who wish to accomplish their unfinished dream of acquiring higher education. When this system is supported with technology particularly with increased use of OERs and Computer mediated learning environment, it is easy on the part of the universities to take the Higher Education even at the grass roots level of the community. Hence, it is evident that OERs and other computer mediated language teaching/learning materials widely available in the web could very well be exploited in the teaching and learning of Tamil as the first language to the aspirants in general and particularly to NRIs of Tamil origin settled elsewhere.

References:

1. Balasubramanian, N and Kadhiravan, S (1999) " Trends in the development of Computer Mediated Instruction", *Journal of Staff and Educational Development International*, 42(2), p.187-183,
2. Creative Commons (2007a). History : " Some Rights Reserved " Building a layer of reasonable copyright. Creative Commons. Retrieved November 17, 2007 from: <http://wiki.creativecommons.org/History>
3. Downes.S. (2007), Models for sustainable open educational resources. *Interdisciplinary Journal of Knowledge and Learning Objects*, 3. Retrieved November 26,2007 from: <http://www.ijkl.org/Volume3/IJKLOv3p029-044Downes.pdf>

4. Jan Hylan (2007) " Open Educational Resources : Opportunities and Challenges ", OECD's Centre for Educational Research and Innovation, Paris, France, from www.oecd.org/edu/ceri
5. MIT(2006). Profiles : From diving to surfing. Retrieved June 7, 2007 from: <http://www.ocw.mit.edu/OcwWeb/Global/AboutOCW/Profiles.htm>
6. Robert.T Rude(K86), Teaching Reading Using Micro Computers. New Jersey: Prentice- Hall, Inc
7. Ronald, M. Kaplan(1980), Lexical Functional Grammar : A Formal System for Grammatical Representations. Cambridge: MIT Press.

இங்கிலாந்தில் தமிழ் கல்வி கற்பிப்பதில் தகவல் தொடர்புத் தொழில் நுட்பத்தின் பாவனை

Using Information and Communication Technology (ICT) in the UK to enhance Tamil Language Teaching

Siva Pillai

Principal Examiner of Cambridge University ASSET Languages-Tamil Language
Chief Examiner London - Edexcel Examination - Tamil Language
Winner of European Award for Languages 2007
Computer Officer, Goldsmiths, University of London, UK
email: pillaisiva@gmail.com

இங்கிலாந்தில் 300,000 மேற்பட்ட தமிழ் மக்கள் குடியிருக்கிறார்கள் என சமீப கணக்கெடுப்புக் காட்டுகிறது. ஐரோப்பிய நாடுகளிலும் கனடா ஐக்கிய அமெரிக்க நாடுகளிலும் குறிப்பிடத்தக்க தொகைகொண்ட தமிழ் மக்கள் குடியிருக்கிறார்கள். வருங்கால தமிழ் குழந்தைகளுக்குத் தமிழ் கற்றுக் கொடுக்கப்படுகிறது. இதில் பெற்றோர்களும் ஊக்கமாக இருக்கிறார்கள்.

இங்கிலாந்தில் 50 மேற்பட்ட சமூகப் பள்ளிகள் தமிழ் மாணவர்களுக்குத் தமிழ் கற்பிப்பதில் (கேட்டல், பேசுதல், வாசித்தல், எழுதுதல்) ஈடுபட்டு வருகின்றன, இச்சமூகப் பள்ளிகள் பெரும்பாலும் சனி ஞாயிற்றுக் கிழமைகளிலும் வாரநாள்களில் பள்ளி முடிந்த பின்பும் நடந்து வருகின்றன, இதில் சில தமிழ் பள்ளிகள் அரசாங்க முன்பள்ளிகளுடன் இணைந்து அரசாங்கத்தால் அங்கீகாரம் பெற்ற தமிழ் மொழிப்பாடத்திட்டத்திற்கு அமையத் தமிழ் மொழியை பயிற்றப்பட்ட தமிழ் ஆசிரியர்களால் கற்பித்து வருகிறார்கள். சில சமூகப் பள்ளிகள் தேசியபாடத்திட்டத்திற்கு அமையவும் பரம்பரை வழிமுறையிலும் குழந்தைகளுக்குத் தமிழ் மொழியைக் கற்பித்து வருகின்றன.

சமீபத்தில் சமூக மொழிகள் (தமிழ் உட்பட) கற்பிப்பதில் பாரிய மாற்றங்கள் அரசாங்க கல்வித்திணைக்கழகத்தால் முன்வைக்கப்பட்டன. இதன் அடிப்படையில் இங்கிலாந்து தேசிய சமூக மொழிகள் கற்பிக்கும் பாடத் திட்டத்தின் கீழ் (National Curriculum for Language Teaching in UK) எல்லா மொழிகளுக்கும் ஒரு பொதுவான பாடத்திட்டம் உருவாக்கப்பட்டது. இதில் தமிழும் ஒரு சமூக மொழியாகக் கருதப்பட்டு அதற்கும் ஒரு பாடத்திட்டக் கட்டமைப்பு இந்நாட்டு எணைய மொழிகளும் சமநிலையில் கற்பிக்கப்பட வேண்டும் என்ற கோட்பாட்டிற்கு ஏற்ப உருவாக்கப்பட்டது. இது பல மொழிகள் கற்கும் ஒரு மாணவன் தான் கற்கும் ஏணைய மொழிகளுடன் தனது சமூக மொழியைப் பார்க்கக் கூடியதாக உள்ளது. இக்கல்வி முறை இங்கிலாந்து பூராவும் நடைமுறைப்படுத்தப்பட்டு வருகிறது. இதன் மூலம் ஒரு மாணவன் அரசாங்க கல்வி பரீட்சை ஸ்தாபனத்தால் (OCR- Oxford Cambridge Royal Society of Arts) ஒரு அங்கீகாரம் பெற்ற சான்றிதழை தான் கற்கும் ஏணைய ஐரோப்பிய மொழிகளுக்கு ஏற்பத் தமிழ் மொழிக்கும் பெறுபேறுகள் ஒவ்வொரு தரத்திலும் கிடைப்பதைக் கண்டு தமிழை ஊக்கத்துடன் மேலும் கற்கிறான்.

துரிதமான தொழில்நுட்ப நடைமுறைகள் கடந்த பதினைந்து வருடங்களில் உலகளாவிய ரீதியில் தமிழ் சமூகத்தில் பல மாற்றங்களை முக்கியமாக இங்கிலாந்தில் ஏற்படுத்தி உள்ளது. தற்காலத்துக்கேற்ற தொழில்நுட்ப வழி முறையில் ஏணைய ஐரோப்பிய மொழிகளைக் கற்பது போல் தமிழைக் கற்றாக மாணவர் விரும்புகிறனர். ஐரோப்பிய மொழிகள் கற்பதில் கற்பிப்பதில் பல காலமாக தகவல்தொழில்நுட்பத்தைக்

கையாண்டு வருகிறார்கள். புதிய தகவல் தொழில் நுட்ப வழி முறை இங்கு பெரிய பங்கை வகிக்கிறது. அன்றாட தொழில் நுட்ப வளர்ச்சிக்கேற்ப தமிழ்மொழியைக் கற்பிக்காத இடத்து மாணவர் தமிழ் கற்கும் ஆர்வம் குன்றிவிடுகிறது. இத்தொழில் நுட்ப வளங்களை எவ்வாறு நடைமுறைப்படுத்தலாம் என்பதை தமிழ் மொழி பாடத்திட்டக் கட்டமைப்பு படிப்படியாக காட்டுகிறது. அதல் முக்கியமாக ஒவ்வொரு தரத்திலும் தொழில் நுட்பத்தை எவ்வாறு மாணவர் மத்தியில் கையாள வேண்டும் என்பதையும் எடுத்துக் காட்டுகிறது. மேலும் இங்கிலாந்து முன்பள்ளிகளுடன் பங்காளித்தத்துவத்தை ஏற்படுத்தி அதற்கு ஆதரவு அளிக்கும் சிரமத்தான நிர்வாகங்களுடன் (உ-ம். Ourlanguages (www.ourlanguages.org), Cilt (National Centre for Languages UK) (www.cilt.org.uk), National resources Centre NRC (www.nrc.org) இணைந்து புதிய தொழில் நுட்ப வழிவகைகளை தமிழ் மொழி கற்பிப்பதில் ஏற்படுத்தப்பட்டிருக்கிறது.

தமிழ் மொழியில் மென்பொருள்களில் ஒன்றான பவர்பொயின்ட் (Powerpoint) பெரிய பங்கு வகிக்கிறது. இதில் ஒளி, ஒலி அசைவு இவை மூன்றையும் ஒன்றிணைத்து மாணவர்களுக்கு ஆர்வம் ஊட்டும் வகையில் தமிழை ஏனைய மொழிகள் போல் கற்பிப்பதற்கு இடம் உண்டு. ஆசிரியர்கள் மாணவர்களை ஈடுபடச் செய்து அவர்கள் மூலம் தமிழ் கற்பிக்கும் வளங்களை அவர்கள் தரத்திற்கு ஏற்ப உருவாக்கலாம். சுயசிந்தனை சுயஆற்றல் சுயஆக்கம் சுயமதிப்பீடு ஆகியனவற்றை மாணவர் மத்தியில் ஏற்படுத்துவதற்கு இது வழி வகுக்கும்.

மென்பொருள் சந்தையில் இவ்வாறான தமிழ் கற்கும் மென்பொருள்களை வாங்க முடியும். இருந்தும் வகுப்பு மாணவர்களால் ஆக்கப்பட்டதற்கு அந்த வகுப்பில் நிறைய ஆதரவையும் ஊக்கத்தையும் ஏனைய மாணவர்கள் இதில் ஈடுபடுபடச் செய்யவும் இது வழி வகுக்கும். அத்தோடு இந்த மென்பொருளை மாணவர்கள் தமது ஏனைய பாடங்களில் அன்றாடம் தமது ஆக்கத்திறனை ஏனையோருக்குக் காட்டுவதற்குப் பயன் படுத்துகிறார்கள். எமது தமிழ் பள்ளி வகுப்பில் மாணவர்கள் தாமே பேசும் கதைப்புத்தகம் ஒன்றை பவர்பொயின்ட் மூலம் ஆக்கினார்கள். ஒரு மாணவன் படங்களைச் சேகரித்தான். ஒரு மாணவன் கதையை எழுதினான். ஒரு மாணவன் கதையை வாசித்து அதைக் கணினியில் பதிவு செய்தான். எல்லோரும் ஒன்று கூடி மின்கைப் புத்தகத்தை உருவாக்கி ஏனையோருக்குக் காண்பித்தார்கள். இது ஒரு கூட்டு முயற்சி. இவ்வழி மூலம் மாணவர்கள் இடையே ஒரு புரிந்துணர்வை, கூட்டு முயற்சியை வளர்க்க கூடியதாக இருக்கும். மேலும் இந்த வளத்தை தமது பள்ளியின் இணையத்தில் இணைப்பதன் மூலம் உலகில் உள்ள எல்லோரும் இணையம் மூலம் தமிழைக் கற்கலாம். அத்தோடு நாட்டுக்கு நாடு உள்ள மாணவர்கள் தமது படைப்புகளைப் ஒருவருக்கு ஒருவர் இணைய மூலம் பரிமாறிக் கொள்ளலாம். ஏனைய ஐரோப்பிய மொழிகளில் இது பொதுவாக உள்ளது.



இங்கிலாந்தில் இன்றைய வகுப்பறைகளில் வழமையாக இருந்த கரும்பலகைகள் இல்லாது IWB- Interactive White Board) இணைந்தியங்கும் மின்வெண்பலகைகள் முக்கிய பங்கெடுக்கின்றன. இதில் பெரிய வடிவில் தெளிவாக எல்லோரும் பார்த்தறியக் கூடிய வகையில் எழுத்துக்கள், படங்கள் ஆகியனவற்றை காண்பிக்கலாம். எல்லாப் பாடங்களிலும் இந்த வளம் உபயோகிக்கப்படுகிறது. மாணவர் ஒவ்வொருவரது படைப்புகளையும் எல்லோரும் (சினிமா தியேட்டரில் பார்ப்பது போல) பார்க்கலாம். இத்தொழில் நுட்பப் பயன்பாடு படைப்புகளைக் காண்பிக்கும் போது மாணவர்களின் கவனத்தை ஈர்க்கச் செய்வதில் பங்களிக்கிறது. அத்தோடு சொற்களை எழுத்துக்களை படங்களை மின்வெண்பலகையில் ஒரு இடத்தில் இருந்து இன்னொரு இடத்திற்கு அதற்கான மின் பேனாவால் அல்லது ஆசிரியர் மாணவர் தமது கையால் தொட்டு அசையச் செய்யலாம். மின்நிறங்களைத் தரும் பேனா மூலம் அல்லது மின் நிறப்பெட்டியில் கையால் நிறங்களைத் தேர்வுசெய்து வெண்பலகையில் எழுதலாம் அல்லது எழுதியதை அழிக்கலாம். கற்பிக்கும் வளங்களில் இவ்வழி முறை ஒரு பாரிய மாற்றத்தை ஏற்படுத்தி உள்ளது.

தற்போது மாணவர்கள் தமது தமிழ் எழுத்துக்களைத் தட்டச்சுச் செய்வதற்குக் கணினியைப் பயன்படுத்துகிறார்கள். இவ்வளவு காலமும் கணினியில் தொடர்பு கொள்வதற்கு ஆங்கல மொழி மட்டுமே பயன்படுத்தப்பட்டு வந்தது. ஆரம்பத்தில் கணினியில் தமிழை கொண்டு வருவதில் தொழில் நுட்பக் குறைபாடுகள் இருந்தன. தமிழ் தட்டச்சுச் செய்வதற்கு யூனிகோர்ட் தமிழ் தொழில் நுட்பம் கணினியில் இருப்பது மிகவும் சுலபமாக உள்ளது. எல்லோரும் படிக்கக் கூடிய வகையில் இன்று யூனிகோர்ட் ஒரு பொது எழுத்துருவாக உள்ளது. இது மாணவர்களுக்கு ஊக்கத்தைக் கொடுக்கிறது. அவர்கள் தமது படைப்பைச் சரிபார்த்துத் திருத்தி அமைப்பதற்கு இத்தொழில் நுட்பம் பெரிதும் உதவுகிறது.

முக்கியமாக பிரயாணத் தகவல் பரிமாறல் தற்போது கணப்பொழுதில் நாட்டுக்கு நாடு பரிமாறப்படுகிறது. டிஜிட்டல் தொடர்பாடல் (Digital communication) இன்று முக்கிய பங்கம் வகிக்கிறது. மாணவர்கள் முன்பள்ளிகளிலேயே வீடியோ கமிரா பாவிக்கும் தொழில் நுட்பத்தைக் கற்றுக் கொடுக்கிறார்கள். வகுப்பறையில் கற்கும் செயல்பாடுகளை கமிராவில் பிடித்து கற்கும் கற்பிக்கும் வளங்களை உருவாக்குவதில் மாணவர்கள் மிக ஆர்வமாக ஈடுபடுகிறார்கள், இதை இங்கலாந்தில் OurLanguages Project - எனும் சிரமதான ஸ்தாபனம் ஒரு மாணவர் திட்டமாக எடுத்து மாணவர்களை ஊக்கத்துடன் ஈடுபடச் செய்கிறார்கள். இதற்கான திட்டத்தில் எமது சமூகப் பள்ளியும் (Tamil Academy of Language & Arts) 2008-09-ல் தேர்ந்து எடுக்கப்பட்டது.

பாரம்பரிய கல்வி கற்கும் கற்பிக்கும் வழிமுறையிலிருந்து இது வேறுபட்ட விதத்தில் மாணவர்கள் தமிழ் மொழியைக் கற்பதற்கு ஊக்கம் கொடுக்கிறது. டிஜிட்டல் வீடியோ கற்றல் கற்பிப்பதில் பல மாற்றங்களை உண்டாக்கிறது என தனது ஆய்வில் பேர்ன் குறிப்பிடுகிறார்.

" Motivate and engage a wider range of pupils than traditional teaching methods, so providing greater access to the curriculum; (Burn et al 2002)

நாட்டுக்கு நாடு இருக்கும் மாணவர்கள் ஒருவருக்கொருவர் செய்திகளை பரிமாறிக் கொள்கிறார்கள். இந்நாட்டில் நடக்கும் தமிழ் வகுப்பை பல மைல்களுக்கு அப்பால் உள்ள தமிழ் வகுப்பு மாணவர்கள் நேரேயே பார்த்துக் கலந்துரையாடும் தொழில் நுட்ப வசதிகள் அதிகமான பள்ளிகளில் இங்கிலாந்தில் உள்ளன. கணினியிலேயே ஒருவருக்கொருவர் பாடங்களை, கற்பதை, செய்திகளை பரிமாற்றம் செய்து கொள்ளலாம். மாணவர்கள் தன்நம்பிக்கையோடு தமது படைப்புகளை இத்தொழில் நுட்பம் ஊடாகப் பரிமாற்றம் செய்து கொண்டு இருக்கிறார்கள்.

பல இணைய தளங்கள் தமிழ் கல்வி கற்பிப்பதற்கு உருவாக்கப்பட்டுள்ளன. முக்கியமாக தமிழ் இணையப் பல்கலைக்கழகம் (Tamil Virtual University (TVU) Started 1999) குறிப்பிடத்தக்கது.

பல இணையத்தளங்கள் தமிழ் கற்பிப்பதற்கு இருந்தாலும் அதிக ஆய்வுகள் மூலம் மேலை நாட்டுக் குழந்தைகளுக்கேற்ப தமிழ் கற்கும் இணையத்தளம் 1997-ல் University of Pennsylvania ' Web Assisted Learning and Teaching of Tamil@ (WALTT) உருவாக்கப்பட்டது. இது தமிழ் மொழியே தெரியாத ஒருவர் தமிழைக் கற்றுக் கொள்வதற்குப் படிப்படியாக ஆசிரியருக்கும் மாணவர்களுக்கும் ஆக்கமான வளங்களைக் கொண்டு உள்ளது. இவ்வளங்களை இங்குள்ள சமூகப் பள்ளிகள் பயன்படுத்திக் கொள்கின்றன.

ICT-(Information and Communication Technology) தகவல் தொடர்புத் தொழில் நுட்பம் குறிப்பிடத் தக்க அளவில் தமிழ் கற்றல் கற்பிப்பதில் இங்கிலாந்திலும் உலகளாவிய ரீதியிலும் கூட பல மாற்றங்களை ஏற்படுத்தி உள்ளது. மாணவர்களுக்குப் புதிய வழிமுறைகளில் கற்பதில் கற்பிப்பதில் ஊக்கத்தை உண்டுபண்ணி உள்ளது. உலகில் உள்ள எல்லா தமிழ் சமூக மக்களுடன் தற்போது தொடர்பை ஏற்படுத்த இது உதவி வருகிறது. முன்னேற்றங்கள் பல இருந்தாலும் கலாச்சாரம் மொழி கலப்பு தவிர்க்க முடியாதது ஒன்றாகும். உலக சந்தையில் தமிழ் மொழியும் தமிழ் கலாச்சாரமும் மாற்றங்களை ஏற்படுத்தும் என்பதை மறுக்க முடியாது. எமது வருங்கால சந்ததினரிடம் தற்காலத்திகேற்ற தொழில் நுட்ப வழியில் தமிழ் மொழியைக் கற்கச் செய்வது எமது கடைமயாகும்.

அறிவாக்கத்தை அடிப்படையாகக் கொண்ட

கற்றல் - கற்பித்தல்

திலகவதி சங்கர்

மூத்த-தமிழாசிரியர்

போவன் உயர்நிலைப் பள்ளி

சிங்கப்பூர்

அறிமுகம்

கல்வியின் அடித்தளமாக அமைவது மொழிக்கல்வியே என்பது உலகாந்த உண்மை. கணிதத்தில் நிபுணத்துவம் பெற்றாலும், அறிவியலில் நிபுணத்துவம் பெற்றாலும், ஒருவரிடம் மொழித்திறம் இல்லையேல் அவரால் அவ்வவத் துறையில் கூட வெற்றி அடைவது சிரமமே. கடந்த சில ஆண்டுகளாகத் தாய்மொழி கற்றல் குறித்துப் பல சர்ச்சைகள் எழுந்துள்ளன. பெற்றோர்களிடையே தமிழ் கடினம், வீட்டில் தமிழ்ப் பழக்கமில்லை, தேர்வுகள் கடினமாகியுள்ளன போன்ற புகார்கள் ஒரு புறமிருக்க, மறுமுனையில் தேசிய நிலைகளில் மாணவர்களது எழுத்து தமிழ்த் தேர்ச்சி-நிலை சீராக இருந்தாலும் தமிழ்மொழிப் பேசுவவர்கள் குறைந்து வருவது இச்சர்ச்சைகளுக்கு முக்கியக் காரணமாகியுள்ளது. நமது கல்வி அமைச்சர் திரு இங் எங் ஹென் 'பத்தாண்டுகள் தாய்மொழியைக் கற்கும் மாணவர்களால் இயல்பாகத் தங்கள் தாய்மொழியைப் பேச இயலாதது குறித்து வருத்தம் தெரிவித்தார். பாடத்திட்டமும் கற்பித்தல் முறைகளும் இன்றைய மாணவர்களது நிலைக்கேற்ப மாற வேண்டும்; அப்போது தான் நம் மாணவர்கள் எதிர்காலச் சவால்களைச் சந்திக்க முடியும்,' என்றும் கூறினார். (தமிழ் முரசு-செப்டம்பர் 2009).

தமிழ்ப் படிக்கும் மாணவர்கள் தமிழை அந்நிய மொழியாகக் கருதுவதற்கு முன் என் மாணவர்களுக்கு தமிழின்பால் பற்றுதலையும் தமிழுணர்வையும் ஊட்ட விரும்பினேன். கணினி சுதேசிகளாக மாறியிருக்கும் மாணவர்களை அவர்கள் வழியில் சென்று வெல்லவேண்டும் எனக் கருதினேன். ஒரு மொழியில் ஆற்றல் பெறுவதற்குத் தகவல் தொழில்நுட்பம் தடையாக இருக்கும் என்ற சித்தாந்தம் கூறியவர்கள்கூட இன்று பல்லாடகங்களின் தாக்கத்தால் கணினி சுதேசிகளாக மாறியிருக்கும் நம் மாணவர்களை மொழியின்பால் ஈர்க்க அவற்றையே கையாண்ட பல வகுப்பறை அனுபவங்களைப் பற்றி ஆய்வு நூல்கள் வழி கேள்வியுற்றேன். என் தமிழ் வகுப்பறையில் அவற்றை, இன்னபிற வளங்களுடன் இணைத்து என் கற்பித்தல் முறைகளுக்கு மெருகூட்டினேன்-வெற்றி கண்டேன். அம்முயற்சியின் வெற்றியை இங்கு பகிர்ந்துகொள்ள முனைந்துள்ளேன்.

பாடத்தொகுதியின் அமைப்பு

இதப்பாடத்தொகுதி கல்வி அமைச்சு வழங்கியுள்ள பாடத்திட்டத்தையும் பாடப்புத்தகங்களையும் தழுவி அமைக்கப்பட்டது. மாணவர்களது நிலைக்கேற்பவும் சூழலுக்கேற்பவும் கருப்பொருளை ஒட்டி அமையப் பெற்றதால் மாணவர்களுக்குப் ஆய்வு செய்யவும் தகவல்கள் திரட்டவும் ஏதுவாக இருந்தது. பாட நூல்களோடு இணைய வழி தகவல்கள் பெற்ற மாணவர்கள், அவற்றை அறிவாக்க அடிப்படையில் பகுத்துப் பயன்பெறுமாறு பல வாய்ப்புகளைப் பெற்றனர்-இவ்வமைப்பு முறையில் அமைந்த தொகுதியின் விவரங்கள் பின்னிணைப்பு 1-இல் உள்ளது(Annex1). உயர்நிலை 1-இல் பயிலும் மாணவர்களுக்கு 12 வாரத் தொகுதியாகப் பாடங்களை அமைத்து நடத்தினேன். 2 தவணைகளாக அதாவது தவணை 1-இல் 6 வாரங்களுக்கும் தவணை 2-இல் 4 வாரங்களுக்கு மொத்தம் 30 மணிநேரத்துக்குப் பாடங்கள் நடந்தேறின. உயர்நிலை 1-பாடப்புத்தகத்தை ஒட்டி என்னைப்பற்றி, குடும்பம், நண்பர்கள், சமூகம், நாடு ஆகிய கருப்பொருள்களே இந்த அறிவாக்கத்தை அடிப்படையாகக் கொண்ட கற்றல் கற்பித்தல் பாடத்தொகுதியில் அமைந்துள்ளன.

சவால்கள்

- நேரமின்மை- சில செயல்முறை உத்திகள் & பள்ளி முடிந்த செயல்படுத்த வேண்டியிருந்தது.
- ஆசிரியரது உழைப்பு அதிகமாக இருந்தது. முன் தயாரிப்பு அவசியம் என்பதால் ஆசிரியர்கள் மாணவர்களுக்குச் சில விவரங்களைச் சேகரிக்கவேண்டியிருந்தது
- சில மாணவர்களிடம் கணினி இணைய இணைப்பு இல்லாதது

வெற்றிகள்

- மாணவர்களை முனைப்புடன் கற்றனர்-பாடம் மிகவும் சுவையாகவும் இருந்தது. ஆசிரியர்-
- மாணவர் உறவு வலுப்பெற்றது.
- மாணவர்கள் மொழித்திறன் சிறப்பாக இருந்தது. பேட்டி காணும்போது அவர்கள் தமிழைக் கட்டாயமாகப் பயன்படுத்த வேண்டியிருந்தது.
- மாணவர்களது படைப்புகளைப் பார்த்த பிற ஆசிரியர்கள் அதிசயித்தனர் (ஆசிரியர் பகிர்வு-பட்டறை)
- மாணவர்கள் சுயமான கற்றலுக்கு அதிக வாய்ப்பு இருந்தது. குழு உணர்வு, கருத்துப் பரிமாற்றம், தோழமை உணர்வு- வலுப்பெற்றது.
- மாணவர்கள் படைப்புத்திறன், எழுத்துநடை, கருத்துவளம்- நல்ல முன்னேற்றம் இருந்தது.

முடிவுரை

எதிர்கால சவால்களை மாணவனுக்கு வகுப்பில் தத்ரூபமான முறையில் செயல்படுத்தி எடுத்தக்காட்டக்கூடிய கற்றல் பணிகள் அவனை மேலும் சாதிக்க வைக்கும். இத்தொகுதியில் இடம்பெற்றுள்ள பாடங்கள் அனைத்தும் மாணவர்களத் தேவைக்கேற்ப அமைக்கப்பட்டது. மாணவர்களது அன்றாட வாழ்க்கையோடு அவை அவனைத் தொடர்புபடுத்துவதால் மாணவர்களுக்குத் தங்கள் கற்றல் பயனுள்ளதாயும், பொருளுள்ளதாயும் அமைந்துவிடுகிறது. அதுவும் கணினி, பல்லூடகங்கள் வழி கற்பதற்கு வாய்ப்பு ஏற்படும் போது தமிழ் தங்களுக்குரிய மொழியாகிவிடுகிறது. தமிழ் மாணவன் ஆழமாகச் சிந்திக்க முற்படுவான். பிரச்சினைகளுக்குத் தீர்வு காண தன்னம்பிக்கையுடன் செயல்படுவான். இவையாவும் இம்முயற்சியில் நான் கண்ட உண்மைகள். தாமாகச் சிந்திக்கக்கூடிய மாணவன் நிச்சயம் வெற்றிபெறுவான். மாணவர்கள் ஏன் கற்கிறோம், எதற்காகக் கற்கிறோம் என்று சிந்திக்க ஆரம்பித்தால் கற்றல் கற்பித்தல் இனிதே அமையும்.

Annex 1

கரு- என்னைப்பற்றி (Me and My Vision)

கற்றல் விளைவு- மாணவர்கள் அடைவுகளை அமைத்தல் பற்றியும் அதற்குரிய வழிமுறைகள் பற்றியும் அறிந்துகொள்வர். பறர் கூறும் நற்கருத்துகளைப் பயன்படுத்தி மனநிலையில் மாற்றத்துக. பல்லாடக வழிகளைப் பயன்படுத்த பயிற்சி பெற்றிருப்பர்

வாரம் 2	தலைப்பு & திறன்கள்	விளக்கம்	அணுகுமுறை பணிகள்	வளங்கள் அறிவாக்க- விளைவுகள்
2 ம.நே	1. என்னைப் பற்றி நானே ..நானா.. கேட்டுணர்தல், சிந்தித்தல், பாகுபடுத்தல், எழுதுதல், படைத்தல்	1 மாணவர்கள் தங்களது எதிர்கால இலட்சியத்தைப் பற்றியும் அதனை அடைவதற்கு செய்துவருவன பற்றியும் தீர்மானங்களை இயற்றி மூடல் பக்கத்தில் ஏற்றுதல். 2. தமக்குத் தேவையான உதவிகளைப் பற்றிய விவரங்களை-படிவம் வழி ஏற்றம் செய்தல் 3. இணைய விதிமுறைகள் அறிதல் & அபாயங்களையும் பற்றிய விழிப்பணர்வு 4. உறவினர், நண்பர்களுக்குத் தங்கள் பக்கங்களுக்கு அழைத்து -கருத்தாடல் .	1. தனிவேலை -என் கனவு & தீர்மானங்கள்-உரையைக் கேட்டு குறிப்பெடுத்துக் கொண்டனர் 2. தம் தீர்மானம் பற்றி மறு பார்வை-சரி படுத்தினர் 3. என் விருப்பம்-படிவம்- பகுத்தல்- தரவரிசைப்படுத்தினர் 4. கருத்துகளைப் பரிசீலனை-பகுத்து மாற்றம் செய்தனர். தமது குறை-நிறைகள்- கட்டுப்பாடுகள்-விளைவுகள் பற்றி நண்பர் & உறவினருடன் விவாதித்தனர் 5. தீர்மானம்-சீரமைத்தனர்	சுயமாகக் கற்றல்- பொறுப்பு Epistemic Agency அறிவாக்கத்துக்குரிய உரையாடல்-Knowledge building discourse பயன்மிக்க- கருவூலங்கள் Constructive use of sources Cultural dimension கலாச்சார நோகரு Moodle portal வலையோலி-Podcasts- என் கனவு சிந்தனைவரைபடம்- மாதிரிப் பக்கங்கள் Newsmaker Audacity என் விருப்பங்கள்- மாதிரிப்பக்கம்

பாட முன்னறிவு; மாணவர்களுக்கு மேற்குறிப்பிட்டவற்றிலும் -மென்பொருள் இயக்கத்திலும் பயிற்சி அளிக்கப்பட்டது.

ஆசிரியர் கருத்துகள்; மாணவர்கள் சுயமாகக் கற்பதற்கும் மாணவரை மையமாகக் கொண்ட கற்றுலுக்கு இப் பாடம் வழிவகுத்தது. வலையோலிப் பகுதி மட்டுமே முழு வகுப்பு வேலையாக இருந்தது. அதன் பின்னர் மாணவர் தனி வேலையில் ஈடுபட்டனர். மாணவர்கள் தங்கள் கற்றலை தாமே வழி நடத்திச் செல்ல வாய்ப்பிருந்தது.

பாடம் 2.2 குடும்பமும் நல்லுறவுகளும்

கற்றல் விளைவுகள்- மாணவர்கள் அழைப்பு முறைகளையும் குடும்ப விழுமங்களைப் பற்றி அறிந்துணர்வர். தங்கள் பாரம்பரியம் பற்றியும் தெரிந்து கொள்வர்.

வாரம்	தலைப்பு & திறன்கள்	விளக்கம்	அணுகுமுறை பணிகள்	வளங்கள் அறிவாக்க-விளைவுகள்
3 ம.நே	1. அடித்தளம்- சிறு குடும்பம் 2. தாத்தா- பாட்டி- உறவினர்கள் சேர்ந்து பெரிய குடும்பம்	1 மாணவர்கள் சிம்சன்ஸ் எனப்படும் படத்தொடரைப் பார்த்தல் 2. குடும்ப உறுப்பினர்கள்தம் பொறுப்புகளை ஆராய்தல். தமது குடும்பங்களுடன் ஒப்பிடுதல். 3. ஒவ்வொருவரிடமுள்ள சிறந்த பண்புகளைப் பட்டியலிடுதல். குடும்பப் பொறுப்புகளுடன் ஒப்பிடுதல். 4. குடும்ப மரம் உருவாக்குதல்-மூதாதையர்? பூர்வீகம்? தகவல் சேகரிப்பு- தாத்தா-பாட்டியுடன் பேட்டி. 5. பாரம்பரிய, பண்பாடு- பழக்க வழக்கம்- தத்தமக்குரியவற்றைக் கண்டறிதல்- கலந்துரையாடல்- உறவினர்- கருத்தாடல்.	1. மாதிரி பக்கங்கள் வழி ஒப்புநோக்கி படத்தொடர் பற்றிய கருத்தெழுதினர் 2. பட்டியலிடுவதன் வழி சிந்தித்து செயல்பட்டு தகுதிவாரியாக அமைத்தனர் 3. குடும்பப் பண்புகள் - பட்டியல் உருவாக்கினர்- பண்பில் சிறந்தவர்களை அறிய சரிபார்ப்புப் பட்டியல் உருவாக்கி இணையாளருடன் ஒப்பிட்டனர். கருத்துப் பரிமாற்றம் 4. தாத்தா-பாட்டியுடன் பேட்டி- எதிர்ப்பார்ப்புகள்-பகிர்ந்து கொண்டனர் 5. பாரம்பரியம்-தேடல்- தகவல்களைச் சேகரிப்பதன் வழி உறவுமுறைகள் அறிந்துகொள்வர்	(Knowledge Building & Convergent Knowledge-expression) அறிவாக்கம் & அறிவார்ந்த கருத்துப் பரிமாற்றம் Create a Family Tree or family charter. குவிநோக்கு அறிவாக்கம் அறிவாக்கத்துக்குரிய உரையாடல்-Knowledge building discourse பயன்மிக்க- கருவூலங்கள் Constructive use of sources Cultural dimension கலாச்சார நோக்கு கற்பனை விரிவாக்கம் உண்மைச்சூழல் Audacity Cartoon clip-The Simpsons Digital cameras Flip Album

ஆசிரியர் முன்னறிவு-பிளவுப்பட்ட குடும்பங்களில் வரும் மாணவர்களைக் கருத்தில் கொள்ளுதல் அவசியம்-மனம் நோகும்படியான பேச்சுக்கு இடம் தராமல் இருப்பதில் ஆசிரியர் சிரத்தையாக இருக்கவேண்டும்.

ஆசிரியர் கருத்து-நான் இப்பாடத்தைச் 2 வருடங்களாக நடத்தி வருகிறேன்-மாணவர்களது முழுக் கவனத்தை ஈர்த்த பாடம் இதுதான். வேறு 2 பள்ளிகளிலும் இப்பாடத்தை அறிமுகம் செய்தேன்-அவ்வாசிரியர்களும் இவ்வறே கருத்து தெரிவித்தனர். மாணவர்களிடையே பிணைப்பை ஏற்படுத்த முடிந்தது. இப்பாடம் மாணவர்களது சுயமான கற்றலை உறுதிச் செய்தது. சிம்சன்ஸ் படக்காட்சி தொடக்கநடவடிக்கையாக இருந்தது மாணவர்களிடையே உற்சாகத்தை ஏற்படுத்தியது.

பாடம் 3-நட்பு-கற்றல் விளைவுகள்-நட்பினால் ஏற்படும் நன்மை தீமைகளை அறிந்து அறிவுப்பூர்வமாகச் செயலெடுவர்

வாரம் இடம்	தலைப்பு & திறன்	விளக்கம்	அணுகுமுறை பணிகள்	வளங்கள் அறிவாக்க-விளைவுகள்
6-1 மணிநேரம் கணினிக் கூடம்	3.1.என்றும் நட்பு- உயிர் நண்பன் கேட்டுணர்தல் & பகுத்துணர்ந்து எழுதுதல் தகவல் பகுத்தல் சிந்தனை- திறன்	1. மாணவர்கள் நட்பினால் தமக்கு ஏற்பட்ட இனிய அனுபவத்தைப் பகிர்தல். (பழைய நட்பு) நண்பர்களுடன் கருத்தாடல்.	1. நட்புப் பற்றி ஒரு படக்காட்சி- பார்த்தனர். 2.மாணவர்கள் நட்பின் சின்னங்கள் பற்றி கூறினர்-அவற்றைச் சிலர் கொண்டு வந்து காட்டினர்-உணர்ச்சிகளைக் கொட்டினர் 3 நட்பு-பற்றிய பக்கத்தை நிரப் பினர் - படக்கதை உருவாக்குதல் -உயிர் நண்பர்கள்(வீட்டுப்பாடம்)	Real Ideas and authentic problemsபன்முனை கருத்துகளால் அமையப்பெற்ற- அறிவாக்கம் Convergent knowledge மூடல்-உலாவருதல் உண்மைச்சூழல்கள் & சிக்கல்கள் கேலிச்சித்திரம்
7-2 மணிநேரம் கணினிக் கூடம்& இல்லம்	3.2 நட்பராய்தல்	1. பதினம வயது நட்பு-கருத்துகளம். வலைப்பூ பயன்படுத்துதல். 2.குழுவேலை- மாணவர்கள்- பாவனை நடிப்பு- சூழல்கள் தருதல் 3.பெற்றோர் - கருத்து-பேட்டி	1.இணையரது உலகமே நட்பு- பெற்றோர் ஆதங்கம்- மாணவராக பாகம் ஏற்பவர்கள் தமது கருத்துகளைத் தெரிவித்தல். பெற்றோராகயிருப்பவர்கள்-தமது சிந்தனைகள்-அச்சங்களைத் தெரிவித்தல். 3. பேட்டியை ஏற்றம் செய்தல் 4. மாணவர்கள் எல்லா கருத்தகளையும் பார்த்தல்-எ.பி.ப- நிரப்புதல்.	பாவனை திறன் -சூழல்கள் பேட்டி-மாதிரி பக்கங்கள் எ.பி.ப=எண்ணப்பிரதிபலிப்புப் படிவம்-Reflection log
8-2 மணிநேரம் கணினிக் கூடம்& இல்லம் .	3.3 சிறந்த நண்பன் யார்	1.இலக்கிய-நட்பு- கதாமாந்தர்கள்- காட்சி 2.நல்லியல்புகள்- கருத்துதிர்ப்பு 3. நட்பாரந்த நாடுகள்-மலேசியா & சிங்கப்பூர்	1.மாணவர்கள் கதாமாந்தரிடம் காணப்பட்ட பண்புகளை அடையாளம் கண்டு சிந்தனை வரைப்படத்தில்நிரப்புதல் 2.1.நல்லியல்புகள்-இணையாளருடன் வரிசைப்படுத்துதல். 2.2. சரிபார்ப்புப் பட்டியல் தயாரித்தல் 2.3 கதாமாந்தர்களது குணநலன்களுடன் ஒப்பிடுதல்.சுயமதிப்பீடு செய்தல் 3. நட்பாரந்த நாடுகள்-தேசிய கல்விச்செய்தி-வரலாறுப் பக்கங்கள் பார்த்து நட்பில்லாததால் கஏற்படக்கூடிய நிலையை இணையப்பக்கம் வழியாக அறிந்துணர்தல். எ.பி.ப-நிரப்புதல்	கர்ணன்,இராமாயணம்,காந்தி யடிகள்,வெனில் நாட்டுப் பெருவணிகள் சரிபார்ப்புப்பட்டியல் மாதிரிப் பக்கம் கருத்துப் பரிமாற்றம்- KB discourse அறிவாக்கம் வளம் பெறுதல்

ஆசிரியர் கருத்து-மாணவர்கள் முழுமையான ஈடுப்பாட்டை என்னால் உணரமுடிந்தது. Facebook,twitter-
பக்கங்களுடன் ஏற்பட்ட இணைப்பு பாடத்துக்கு மேலும் சுவை கூட்டியது. தனிப்பட்ட,ரகசியமான தகவல்கள்-
பொதுமக்களின் பார்வைக்கு வர்வது குறித்து மாணவர்களிடம் விவாதிக்க வேண்டியிருந்தது-மாணவர்கள்
தனிப்பட்ட ரகசியமான தகவல்கள் ஏற்றம் செய்வது -கவனத்தில் கொண்டனர். எனக்கு பிடித்த பாடமாக
இருந்தது. மாணவர்களிடம் நான் பலவற்றையும் கற்றுக்கொண்டேன்.

பல்லாடகப் பயன்பாட்டில் தமிழ்ப் பாடத்திட்டம்

(ஆக்கமும் பயனும்)

முனைவர் ஆ. மணவழகன்

தமிழ்த்துறைத் தலைவர்

அறிவியல் மற்றும் மானுடவியல் புலம்

எஸ்.ஆர்.எம். பல்கலைக்கழகம்

காட்டாங்குளத்தூர் 603 203.

காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு.

Email: Tamilmano77@yahoo.com

எந்த ஒரு மொழியின் புற வடிவமும் அதன் அகப் பொருண்மைகளும் காலந்தோறும் மாற்றத்திற்கு உட்பட்டே வருகின்றன. இம்மாற்றத்தில் அம்மொழி பேசுவோர் கையாளும் தொழில்நுட்பங்களும், புதிய கண்டுபிடிப்புகளும் பெரும்பங்கு வகிக்கின்றன. தொழில்நுட்பங்களுக்கு இயைந்து, அதனோடு சேர்ந்து தம்மையும் தகவமைத்துக்கொள்ளாத எந்தவொரு மொழியும் காலப்போக்கில் மக்கள் பயன்பாட்டிலிருந்து அந்நியப்படுத்தப்படும் சூழலுக்கு ஆளாகிறது. உலகின் எந்த மொழியும் இதற்கு விதிவிலக்கல்ல. அதேவேளையில், தமிழோடு தோன்றிய அல்லது சம காலத்தில் புழங்கிய எத்தனையோ மொழிகள் இன்று பயன்பாடு அற்றுப்போக அல்லது வேறு ஒன்றாகத் திரிந்துபோக, தமிழ் மட்டுமே காலத்தைப் புறந்தள்ளி என்றும் சீரிளமையோடுத் திகழ்கிறது. இதற்கு, தமிழ் நவீன கண்டுபிடிப்புகளைத் தன்னுள்ளே ஏற்று அதற்குத் தக தன்னைத் தகவமைத்துக்கொண்டதோடு, தானும் அவற்றோடு சேர்ந்து வளர்ந்ததே காரணம் எனலாம்.

ஓலைச் சுவடிகளின் வழியாகப் பயணித்த தமிழ், இன்று கணினிவழித் தன் பயணத்தை இனிதே மேற்கொண்டு வருகிறது. இடைப்பட்ட காலகட்டங்களில் தமிழ் மொழியின் அகப்-புற வடிவங்கள் பல்வேறு மாறுதலுக்கு உட்பட்டே வந்திருக்கின்றன. இலக்கியப் பொருண்மைகளும் காலத்திற்கும் தேவைக்கும் ஏற்ப தம்மை மீள் ஆய்விற்கு உட்படுத்தியே வளர்ந்திருக்கின்றன/வளர்கின்றன. சங்க இலக்கியம் தொடங்கி, அற இலக்கியம், காப்பியம், பக்தி இலக்கியம், சிற்றிலக்கியம், உரைநடை இலக்கியம் என்று பல வடிவங்களோடு சூழலுக்கு ஏற்ப தம்மை வெளிப்படுத்திக் கொண்டது தமிழ். இந்நிலையில், சென்ற நூற்றாண்டின் பிற்பகுதி தொடங்கி தமிழ் மொழியின் அகப்பொருண்மைகளைப் புதுக்கவிதை, சிறுகதை, புதினம், கட்டுரை போன்ற இலக்கிய வடிவங்கள் ஆட்கொண்டு வருகின்றன. நவீன இலக்கிய வடிவங்களான SVRஇவை இன்று தமிழிற்கு உகந்ததாக உள்ளன. மேலும், இவ்வடிவங்கள் அச்சுவடிவில் மட்டுமின்றி, கணினிப் போன்ற தொழில்நுட்பங்களையும் தங்களுக்குச் சாதகமாக்கியுள்ளன. எளிய வடிவில் மொழியை, மொழியின் இலக்கியங்களைப் பரவலாக்க, பாதுகாக்க, அனைவருக்கும் எட்டச்செய்ய இத்தொழில்நுட்பப் பயன்பாடு பெரிதும் தேவையாக உள்ளது. அவ்வகையில், பாடத்திட்டத் தமிழைப் பல்லாடகப் பயன்பாட்டில் வடித்துக் கொடுத்தலும், தமிழின் இலக்கிய வடிவங்களை (சங்க இலக்கியம் முதல் தற்கால இலக்கியம் வரை) பல்லாடகத்தின் வழி எளிதாக்கி, அனைவருக்கும் எட்டச்செய்தலும் இன்றைய முக்கியத் தேவைகளாக உள்ளன. காட்சி வடிவிலும், அசைவூட்டங்களுடனும் (Animation) உருவாக்கப்படும் இவ்வகைக் கணினித்தமிழ் காலத்தை வென்றதாக, நீண்டகாலப் பயனைக்கொண்டதாக அமையும் என்பது உறுதி.

இந்நோக்கங்களின் அடிப்படையில், *சொல்லோவியம்*, *காந்தள்*, *உயிரோவியம்* என்ற மூன்று முன்வரைவு 'தமிழ்ப் பல்லாடகத் தொகுப்புகள்' உருவாக்கப்பட்டுள்ளன. அவற்றின் ஆக்கம், அமைப்பு, பயன்பாடு குறித்து இக்கட்டுரை விளக்குகிறது.

1 சொல்லோவியம்

சொல்லோவியம் என்பது சிறுவர்களுக்கான '**படவிளக்க அகராதி**' (Picture Dictionary). தமிழின் அன்றாடப் பயன்பாட்டுச் சொற்களை எழுத்து, சொல், படம், படவிளக்கம், சொற்பயன்பாடு என்ற வகையில் இது அறிமுகப்படுத்துகிறது.

அமைப்பு

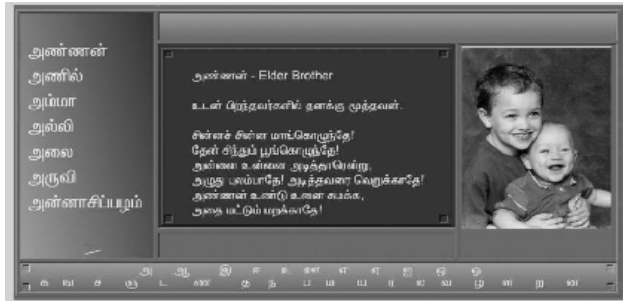
சிறுவர்களுக்குத் தமிழ் எழுத்துகளையும், சொற்களையும் மனத்தில் பதியும் வண்ணம் அறிமுகப்படுத்தும் தொகுப்பு இது. தமிழ்ச் சொற்களை வெறும் சொற்களாக மட்டும் அறிமுகப்படுத்தாமல், சிறுவர்கள் விரும்பிப் பார்க்கும் வகையில் படங்களோடும் ஒலி வடிவமும் உருவாக்கப்பட்டுள்ளது. அகர வரிசையின் அடிப்படையில், பெயர் மற்றும் வினைச் சொற்கள் 300 இடம்பெற்றுள்ளன.

சொல் (பெயர், வினை) – இணையான ஆங்கிலச்சொல் – விளக்கம் - சொற்பயன்பாடு (கவிதை வடிவம்) – சொல்லிற்கு ஏற்ற படம் – ஒலி வடிவம்

என்ற அமைப்பில் இது உருவாக்கப்பட்டுள்ளது.

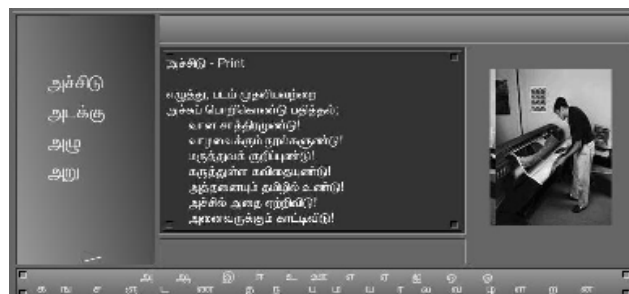
பெயர்

எழுத்துகளின் அகர வரிசையின் அடிப்படையில், ஒவ்வொரு எழுத்திற்குமான சில பயன்பாட்டுப் பெயர்ச்சொற்கள் கொடுக்கப்பட்டுள்ளன. ஒவ்வொரு சொல்லுக்கும் இணையான ஆங்கிலச் சொல், சொல்லுக்கான விளக்கம், சொற்பயன்பாட்டுக் கவிதை, சொல்லிற்கு ஏற்ற படம், கொடுக்கப்பட்டுள்ள செய்திகள் முழுமைக்குமான ஒலி வடிவம் ஆகியன கொடுக்கப்பட்டுள்ளன.



வினை

பெயர்ச்சொற்களைப் போன்றே அகர வரிசையின் அடிப்படையில் ஒவ்வொரு எழுத்திற்குமான சில வினைச்சொற்கள் படவிளக்கத்தோடு கொடுக்கப்பட்டுள்ளன. இதுவும் மேற்கண்ட தன்மைகளைக் கொண்டுள்ளது.



பயன்பாடு

இந்திய மற்றும் வெளிநாட்டுவாழ் தமிழ்க் குழந்தைகளுக்கு எளிய முறையில் தமிழ் எழுத்துகள், சொற்கள், சொற்பொருள் ஆகியவற்றை அறிமுகப்படுத்த, பயிற்றுவிக்க இத்தொகுப்பு பெரிதும் பயன்படும். குறிப்பாக, பொருட்களை(உருவம்) நேரடியாகப் பார்த்து, அவற்றைப் பற்றி அறிந்துகொள்ளும் வாய்ப்பினைப் பெறாத தமிழ்க் குழந்தைகள் இத்தொகுப்பின் மூலம் பெரிதும் பயன்பெறுவர். இது ஒரு முன் உருவாக்கத் திட்டம். இதில் சொற்பயன்பாட்டைத் தேவைக்கு ஏற்ப அதிகப்படுத்திக்கொள்ளலாம். மழலையர் கல்வி மற்றும் தொடக்கக் கல்வி மாணவர்களுக்கு ஏற்ற வகையில் இது உருவாக்கப்பட்டுள்ளது. இதனை உருவாக்க 'ப்ளாஷ்' பல்லாடகம் பயன்படுத்தப்பட்டுள்ளது.

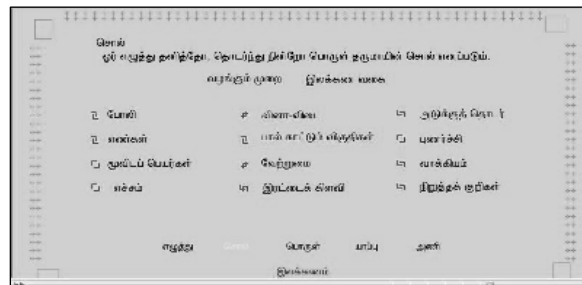
காந்தள்

தமிழ் மொழிக்கான 'மின்கையேடு' (e-guide for Tamil) காந்தள் எனப்படும் பல்லாடகத் தொகுப்பு. இதுவும் 'ப்ளாஷ்' பல்லாடகத்தைப் பயன்படுத்தி உருவாக்கப்பட்டுள்ளது. இத்தொகுப்பின் முகப்புப் பக்கத்தில் அறிமுகம், இலக்கியம், இலக்கணம், பிற என்ற நான்கு முதன்மை இணைப்புகள்(link) கொடுக்கப்பட்டுள்ளன. ஒவ்வொரு இணைப்பின் வழியாகத் திறக்கப்படும் பக்கங்களிலும் உள்ளீடுகளுக்கு (Content) ஏற்ப பல துணை இணைப்புகள் கொடுக்கப்பட்டு, தகவல்கள் விளக்கப்பட்டுள்ளன.



அமைப்பு

- முதன்மை இணைப்புகளுள் ஒன்றான 'அறிமுகம்' என்ற இணைப்பில் இப்பல்லாடகத்திற்கான அறிமுகத்தோடு, தமிழ் மொழி வரலாறு, தமிழ் மொழி பேசுவோர் பற்றிய குறிப்புகள் கொடுக்கப்பட்டுள்ளன.
- 'இலக்கியம்' என்ற முதன்மை இணைப்பில், சங்க இலக்கியம், காப்பியம், பக்தி இலக்கியம், சிற்றிலக்கியம், தற்கால இலக்கியம் ஆகியவை தனித்தனி துணை இணைப்புகளின் வழி அறிமுகப்படுத்தப்பட்டுள்ளன.



- 'இலக்கணம்' என்ற இணைப்பு, தமிழின் ஐந்திலக்கணங்களான எழுத்து, சொல், பொருள், யாப்பு, அணி என்பவற்றிற்கான தனித்தனி இணைப்புகளைக் கொண்டுள்ளது. இவற்றின் வழி தமிழின் ஐந்திலக்கணங்கள் பற்றிய குறிப்புகள் எளிய விளக்கத்தோடும் எடுத்துக்காட்டுகளோடும் அறிமுகப்படுத்தப்படுகின்றன.

- முதன்மை இணைப்பில் நான்காவதாக உள்ள 'பிற' என்ற இணைப்பானது, தமிழின் சிறப்புகள், தமிழரின் சிறப்புகள் ஆகியவற்றை உணர்த்தும் வகையில், தமிழர் அளவை முறைகள், தமிழர் எண்கள், ஆயக் கலைகள், கிழமைகள், மாதங்கள் போன்றவற்றிற்கான தனித்தனி துணை இணைப்புகளைக் கொண்டுள்ளது.

பயன்பாடு

தமிழின் எழுத்துகளைப் படிக்கத் தெரிந்த ஒருவர், அடுத்த நிலையில், தமிழைத் தவறின்றி எழுதுவதற்கான, சொற்றொடர்களை அமைப்பதற்கான அடிப்படை இலக்கணங்களை அறிந்துகொள்ளவும், தமிழ் மொழியின் வரலாறு பற்றியும், அதன் சிறப்பு பற்றியும் அறிமுகப்படுத்திக்கொள்ளவும் இத்தொகுப்பு பெரிதும் உதவும். படவிளக்க அகராதியைத் தமிழ் எழுத்துகளை, வார்த்தைகளைப் பயில்வதற்கான முதல் நிலையாகக் (நிலை-I) கொண்டால் அதன் அடுத்த நிலை (நிலை-II) இத்தொகுப்பு எனலாம். மேலும், தமிழர்களுக்கே உரித்தான அடையாளங்களான தமிழ் எண்கள், தமிழர் அளவு முறைகள், ஆயக் கலைகள் போன்ற சில கூறுகளையும் கொண்டுள்ளது இதன் சிறப்பு. பள்ளிக்கல்வியில் இடைநிலை மாணவர்களுக்கு இத்தொகுப்பு மிகவும் பயனுள்ள ஒன்று.

2 உயிரோவியம்

பழந்தமிழர் வாழ்வியல் முறைகளை சங்க இலக்கியக் காட்சிகளைக் கொண்டு விளக்கும் பல்லாடகத் தொகுப்பு இது. பழந்தமிழரின் புறவாழ்வு வீரம், கொடை, விருந்தோம்பல், நட்பு என்று விரிந்து செல்வது. அகவாழ்வு அன்பால் கட்டமைக்கப்பட்டது. சங்க காலத்தில் அகமும் புறமும் இரு கண்களென்ப போற்றப்பட்டன. அகவாழ்வு களவு, கற்பு என்ற இரு நிலைகளைக் கொண்டது. தமிழர்களின் இவ்வகை அக, புற வாழ்வியை சங்க இலக்கியப் பாடல்களைக் கொண்டு காட்சிகளாக்குகிறது இத்தொகுப்பு.

அமைப்பு

இதில், அகம், புறப் பொருண்மைகளுக்கும் அதன் உட்பகுப்புகளுக்கும் ஏற்ப தேர்ந்தெடுக்கப்பட்ட சங்க இலக்கியக் பாடல்கள் காட்சிகளாக்கப்பட்டுள்ளன. காட்சிக்கேற்ற மூலப் பாடல், பொருள், கவிதை வடிவில் சூழல் விளக்கம், பாடல் பொருண்மையை விளக்கும் ஓவியம் என்ற அமைப்பில் இது உள்ளது.

அகம்

அகம் என்ற பகுப்பு களவு, கற்பு என்ற இருவகைப் பிரிவுகளையும், அதில் சுமார் 20 எடுத்துக்காட்டுப் பாடல்களையும் கொண்டது.

களவு

அகப்பகுதியின் ஒரு கூறான களவு என்பது, முதல் சந்திப்பு, செம்புலப் பெயல் நீர், உச்சிவெயிலில் உருகும் வெண்ணை, துயரத்தில் துணை நிற்போன், காணாததால் கலக்கம், வரைவு கடாவுதல், காத்திருந்து கண்ணீர் மல்கி, அறத்தொடு நிறற்றல், வரைபொருள் பிரிவு, உடன் போக்கு, தோழியின் உள்ளம், செவிலி தேடுதல், கண்டோர் கூற்று ஆகிய தனித்தனித் தலைப்புகளைக் கொண்டுள்ளது. இவற்றின்வழிப் பழந்தமிழரின் களவு வாழ்க்கை நிகழ்வுகள் காட்சிப்படுத்தப்பட்டுள்ளன.



கற்பு

அகப்பகுதியின் மற்றொரு கூறான கற்பு என்பது, அறியா பெண்ணும் அகமகிழ்ந்த செவிலியும், வினையே ஆடவர்க்கு உயிரே, தனிமையில் தலைவி, கார்காலம் கண்டு கலங்குதல், தலைவியின் வருத்தத்திற்குக் காரணம், வந்தான் தலைவன், பாகனுக்கு நன்றி ஆகிய தனித்தனித் தலைப்புகளைக் கொண்டுள்ளது. இவற்றின்கீழ் பழந்தமிழரின் கற்பு வாழ்வியல் நிகழ்வுகள் காட்சிப்படுத்தப்பட்டுள்ளன.

புறம்

‘புறம்’ என்பது தனிப்பகுதி. இதில், தமிழுக்கு மரியாதை, உயிர் சிறிது! மானமோ பெரிது!, விருந்தோம்பல், கொடை, வஞ்சினம் கூறுதல், நட்புக்கு மரியாதை, இலக்கியத்தில் பெண்கள், குழந்தையே செல்வம், கல்வியின் பெருமை, அறிவுறுத்தல், நிலையாமை ஆகிய தனித்தனித் தலைப்புகள் இடம்பெற்றுள்ளன. இவற்றின்கீழ் சுமார் பதினேழு எடுத்துக்காட்டுப் பாடல்கள் கொடுக்கப்பட்டுள்ளன. இவற்றின்வழி பழந்தமிழரின் புறவாழ்க்கையின் பல்வேறு சிறப்புக்கூறுகள் காட்சிப்படுத்தப்பட்டுள்ளன.



பயன்பாடு

பழந்தமிழரின் அகப்புற வாழ்க்கையின் மேன்மையைத் தமிழ் மாணவர்களுக்கும் சங்க இலக்கியம் கற்க வாய்ப்பற்ற உலகத் தமிழர்க்கும் அறிமுகப்படுத்துவதாக இத்தொகுப்பு உள்ளது. சங்க இலக்கியத்தை இவ்வகை எளிய முறையில் அறிமுகப்படுத்துவதன் மூலம், சங்க இலக்கியங்களின் சிறப்புகளை இளைய தலைமுறைக்கு உணர்த்துவதோடு, அவ்விலக்கியங்களின் மீதான நாட்டத்தை அதிகப்படுத்தி அவற்றை முழுமையாக விரும்பிப் படிக்கச் செய்யவும் இத்தொகுப்பு வழிவகுக்கிறது. மேலும், கல்விப் புலங்களில் சங்க இலக்கியப் பாடத்திட்டங்களை மாணவர் விரும்பும் வண்ணம் அமைத்துக்கொடுக்க இத்தொகுப்பு ஒரு முன்னோடித்திட்டமாக அமைகிறது.

Computer Aided Learning in Tamil Sentences

Dr.G.Singaravelu,

Reader,UGC-Academic Staff College,

Bharathiar University,Coimbatore-641 046.Tamilnadu

Singaravelu.bu@gmail.com

Introduction

Sentence has unique place in acquiring knowledge of any language and it is a backbone of the language. Learning to make a sentence in Tamil Language is indispensable for better communication. Learning sentence pattern in Tamil is easy but it is difficult to the young learners to acquire the sufficient sentence patterns in Tamil. Present methods of learning Sentences in Tamil are not effective to the young learners in improving their communicative competencies in Tamil. Challenging innovative Computer Aided Learning can be supported to the young learners to learn more sentence patterns for suitable communicative transactions of oral as well as written in Tamil. The researcher endeavoured to prepare a computed aided learning package for acquiring more sentence patterns in Tamil for the young learners at standard IV. . The study enlightens the effectiveness of Computer Aided Learning in Sentences of Tamil at standard IV

Objectives of the study: 1.To find out the problems of conventional methods in learning sentence pattern in Tamil.2.To find out the significant difference in achievement mean score between the pre test of control group and the post test of control group.3.To find out the significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group.4.To find out the impact of Computer Aided Learning in Sentences of Tamil at standard IV.

Hypotheses of the study:1.Learners of standard IV have problems in learning sentences in Tamil.2.There is no significant difference in achievement mean score between the pre test of control group and the post test of control group. 3. There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group. 4. Computer Aided Learning is more effective than conventional methods in Learning Tamil sentence at standard IV.

Variables

The independent variables namely Computer Aided Learning and the dependent variable namely achievement test score were used in this study.

Delimitations of the Study

The responsibility of the researcher is to see that the study is conducted with maximum care in order to be reliable. However, the following delimitations could not be avoided in the present study.1. The study is confined to 80 students of standard IV studying in Panchayat Union Primary school, Vadavalli, Coimbatore. 2 .The study is confined to learning Tamil sentence only.

Methodology: Equivalent group experimental method was adopted in the study. **Sample:** Eighty pupils of studying in standard IV from Panchayat Union Primary school, Vadavalli, Coimbatore were selected as sample for the study. Forty students were considered as Controlled group and another forty were considered as Experimental group. **Tool:** Researcher's self-made achievement test was used as a tool for the study. An achievement test was framed for making sentences in different patterns.

Construction of tool:

The investigator's self made Achievement test was used for the pretests and post tests of both control groups and experimental groups. The same question was used for both pre and post tests to evaluate the pupils' skills of sentences framing in Tamil through objective types of question which carried one mark for each question and contained 100 marks.

Pilot study

In order to ascertain the feasibility of the proposed research and also the adequacy of the proposed tools for the study a pilot study had been undertaken. During the pilot study, the problem under study had been finely tuned. Sufficient number of model question papers were prepared and distributed to 10 students of standard IV in Panchayat Union Primary school, Vadavalli, Coimbatore for the pilot study. This exercise was repeated twice over two sets of 10 students each. The clarification raised by the students was cleared then and there and the filled answer scripts were collected by the researcher. These students were selected in such a way that they were not part of either the control group or experimental group.

Reliability of the tool

Reliability had been computed using test-retest method and the calculated value is 0.87. The value is quite significant and implies that the tools adopted were reliable. Hence the reliability was established for the study.

Validity of the tool

Subject experts and experienced teachers were requested to analyse the tool. Their opinions indicated that the tool had content validity.

Procedure of the study: 1. Identification of the problem by administering pre-test to the both groups. 2. Planning. 3. Preproduction, production and post-production of CAL. 4. Treatment .5. Administering the post-test.

Data collection:

The researcher administered pretest to the pupils with the help of a teacher and Headmaster. The question papers were given to the individual learners and evaluated learning obstacles of the learners were identified by the pretest. The causes of low achievement by unsuitable methods were found out. Computer Aided Learning was used in the classroom for learning Tamil sentences for one week. The posttest was administered and the effectiveness of the Computer Aided Learning was assessed.

Data analysis

Statistical technique test was computed for the study.

TESTING OF HYPOTHESES

Hypothesis 1:

Pupils of standard IV have obstacles in learning Sentences in Tamil at Panchayat Union Primary school, Vadavalli,,Coimbatore.

In the pre-test, Pupils score 23% marks in learning Tamil sentence through conventional method and the Experimental group students score 77% marks. It shows that Pupils of standard IV have problems in learning Sentences in Tamil at Panchayat Union Primary school, Vadavalli,Coimbatore.

Hypothesis 2:

There is no significant difference between the pret test of control group and post test of control group in achievement mean scores of the pupils in learning Sentences in Tamil at standard IV in Panchayat Union Primary school, Vadavalli,,Coimbatore.

Table -1

Stages	N	Mean	S.D.	df	t- value	Result
Pretest control group	40	12.58	2.90	78	0.17	Insignificant at 0.05 level
Post test control group	40	12.70	3.12			

The table showing achievement mean scores between pre test of control group and posttest of Control group.

The calculated " t ' value is (0.17) less than table value (1.99). Hence null hypothesis is accepted at 0.05 levels. Hence there is no significant difference between the pre test of control group and post test of control group in achievement mean scores of the learners in learning sentences in Tamil.

Hypothesis 3:

There is no significant difference between the pre test of Experimental group and post test of Experimental group in achievement mean scores of the pupils in learning Sentences in Tamil.

Table-2

Stages	N	Mean	S.D.	df	t- value	Result
Pretest Experimental group	40	13.70	3.24	78	8.67	Significant at 0.05 level
Post test Experimental group	40	19.65	3.21			

The table showing achievement mean scores between pretest of Experimental group and posttest of Experimental group.

The calculated 't' value is (8.67) greater than table value (1.99). Hence null hypothesis is rejected at 0.05 level. Hence there is significant difference between the pre test of Experimental group and post test experimental group in achievement mean scores of the learners of Tamil sentences.

Hypothesis 4.

Computer Aided Learning is more effective than existing methods in learning sentences in Tamil at standard IV.

Achievement mean scores of the learners in post-test of control group is 12.70 and the achievement mean scores of the learners post test of Experimental group is 19.65. Score of the post test of Experimental group (19.65) is greater than Pre test of Experimental group (13.70). It shows that learning sentences by using Computer Aided Learning is more effective than conventional methods.

Findings:

1. In the pre-test, Pupils score 23% marks in learning Tamil sentences through conventional method and the Experimental group students score 77% marks. It shows that Pupils of standard IV have problems in learning Sentences in Tamil at Panchayat Union Primary school, Vadavalli, Coimbatore.
2. There is no significant difference between the pre test of control group and post test control group in achievement mean scores of the pupil of standard IV in learning Tamil sentences through Computer Aided Learning at Panchayat Union Primary school, Vadavalli, Coimbatore.
3. There is significant difference between the pre test of Experimental group and
4. post test of Experimental group in achievement mean scores of the pupils in
i. learning Tamil sentences.
5. Computer Aided Learning is more effective than existing methods in learning
i. sentences in Tamil at standard IV.

Educational Implications

1. Computer Aided Learning in Tamil can be extended to primary level, secondary level and higher secondary level.
2. It can be encouraged to implement to use in adult education
3. It may be activated in teachers education
4. It may be implemented in alternative school
5. Eliminating the problems of slow learners by using it
6. It may be more supportive to promote Sarva Siksha Abiyan in grass root level.

Conclusion

The study reveals that Students of standard IV in Panchayat Union Primary school, Vadavalli,,Coimbatore have problems in learning Tamil sentences through conventional method. Learning sentences in Tamil through Computer Aided Learning is more effective than conventional methods. Hence it will be more supportive to enrich sentences in Tamil at primary education.

References

1. **Ray, William, S.**(1960) "An introduction of experimental design. The Macmillan company: New York.
2. **Ravichandran.T, Computer Assisted Language Learning,** (Paper presented and published in the Proceedings: National Seminar on CALL, Anna University, Chennai, 10-12 Feb. 2000, pp. 82-89.)
3. **Vasu Renganathan(2009)**Enhancing the Process of LearningTamil with Synchronised Media,Tamil internet conference,INFITT:Germany
4. **Singaravelu.G, (2009)**Effectiveness of Multimedia Package in Learning Vocabulary in Tamil Tamil internet conference,INFITT:Germany
5. **Sampath.K,Paneerselvam.A and Santhanam.S(1998),** Introduction to Educational Technology,Sterling publishers Pvt Ltd.

கணினிவழி மாணவர்களிடத்தில்

தமிழ்மொழியை வளர்த்தல்

திருமதி சிவகௌரி கலியமூர்த்தி

பீக்கன் தொடக்கப்பள்ளி

பின்னணி - முன்னுரை

சிங்கையில் தமிழ் பேசும் குடும்பச்சூழலிலிருந்து வரும் மாணவர்களின் எண்ணிக்கை குறைந்து வருகிறது என்று கல்வி அமைச்சின் தமிழ்மொழி மறு ஆய்வுக்குழுவின் (2005) புள்ளி விவரங்கள் காட்டுகின்றன. தொடக்கநிலை உயர்நிலை பள்ளிகளில் தமிழ்மொழியைக் கற்றவர்கள் அம்மொழியைக் கருத்துப் பரிமாற்றத்துக்குப் பயன்படுத்துவதில் தயக்கம் காட்டுகின்றனர் என அவ்வாய்வு சுட்டிக்காட்டுகிறது. ஆய்வுக்குழு கல்வி அமைச்சுக்கு வழங்கிய பரிந்துரைகளுள் முனைப்புமிக்க கல்வித்திட்டத்தை உருவாக்குவதும் ஒன்றாகும். அதன் அடிப்படையில் புதிய கல்வித்திட்டம் பேச்சுத்திறனுக்கு முதன்மை அளிக்கும்வண்ணம் தமிழ்மொழிப்பாடத்திட்டத்தை (2008-ஐ) வடிவமைத்துள்ளது. இரட்டை வழக்குள்ள தமிழ்மொழியில் பேசுதலுக்குப் பேச்சுத்தமிழை மாணவர்கள் பயன்படுத்துதல் வேண்டும் என்றும் பரிந்துரைக்கப்பட்டுள்ளது.

(Digital natives) கணினிச் சூதேசிகள் என்று அழைக்கப்படும் இன்றைய மாணவர்கள் இளம் வயதிலேயே கணினியை இயக்குவதில் வல்லுநர்களாக விளங்குகின்றனர். ஆதலால், புதிய தமிழோசை_ பாடப்புத்தகங்களிலிருந்து கணினிவழிக் கற்றலுக்கு ஏற்ற பாடங்கள் தேர்ந்தெடுக்கப்பட்டன. கணினிவழிக் கதைகூறுதல், சக மாணவர்களுடன் ஒன்றிணைந்து கற்றல், முதலிய கூடிக்கற்றல் (cooperative strategies) உத்திகளைப் பயன்படுத்திப் பாடங்கள் தயாரிக்கப்பட்டன. பெற்றோரையும் நடவடிக்கைகளில் ஈடுபடுத்தினால் மாணவர்களின் தமிழ்ப்பழக்கம் அவர்களுடைய இல்லங்களிலும் வேரூன்றும் என்ற நோக்கத்தில் இப்பாடங்கள் தயாரிக்கப்பட்டன.

பீக்கன் தொடக்கப்பள்ளி, FutureSchools@Singapore எதிர்காலப்பள்ளிச் செயல்திட்டத்தில் இணைந்துள்ளது. இங்கு ஏழு தொடக்கநிலை மூன்றாம் வகுப்பு மாணவர்கள், பனிரெண்டு தொடக்கநிலை இரண்டாம் வகுப்பு மாணவர்கள் மற்றும் பத்து தொடக்கநிலை ஒன்றாம் வகுப்பு மாணவர்கள் தமிழ்மொழியைக் கற்கிறார்கள். இங்கு பயிலும் மாணவர்களிடம் எத்தகைய முன் கணினி இயக்கத்திறன்களும் எதிர்பார்க்கப்படவில்லை. ஆதலால், மாணவர்களுக்குத் தேவைப்படும் அடிப்படைக் கணினி இயக்கத்திறன்களைக் கருத்திற்கொண்டு பாடங்கள் தயாரிக்கப்பட்டன. தேவைக்கேற்ப அவ்வப்போது கணினி இயக்கத்திறன்களும் படிப்படியாகக் கற்பிக்கப்பட்டன. அனைத்து மாணவர்களும் கணினிவழிக் கற்றலை மேற்கொள்ளும் வகையில் பாடங்கள் தயாரிக்கப்பட்டன. தமிழை இரண்டாம் மொழியாகப் பயிலும் தொடக்கநிலை ஒன்றாம் வகுப்பு மாணவர்கள் கணினியைப் பயன்படுத்திப் தமிழ்மொழியைக் கற்பதில் எப்படி மேம்பட்டனர் என்பதை விளக்குவதே இக்கட்டுரையின் நோக்கமாகும். இதில் கண்டறியப்பட்ட முடிவுகளும் பரிந்துரைக்கப்படும் நடவடிக்கைகளும் மற்ற பள்ளிகளிலும் நடைமுறைப்படுத்த ஏதுவாக அமையும் என்று நம்பப்படுகிறது.

இவ்வாய்வை மேற்கொள்வதற்கான காரணம் :

1. மாணவர்களுள் பலர் தற்போது தமிழ்ப்பழக்கம் இல்லாத வீடுகளிலிருந்து தொடக்கநிலைப் பள்ளியில் அடியெடுத்து வைக்கிறார்கள். இவர்களை எப்படித் தமிழ்மொழியின்பால் ஈர்ப்பது? இவ்வினாவுக்கு விடை கண்டது இக்கட்டுரை.

கட்டுரைச் சுருக்கம்

தமிழ்மொழிப் பாடக்கலைத்திட்டத்தை மையமாகக் கொண்டு பாடத்திட்டம் வடிவமைக்கப்பட்டது. கணினி வளங்களும் ஏனைய பயிற்றுக்கருவிகளும் உருவாக்கப்பட்டன. பின், மாணவர்கள் கணினிவழி தமிழ்மொழி கற்றல் பயணத்தை மேற்கொண்டு மேம்பாடு கண்டனர்.

பொதுநோக்கம்

கணினிவழிக் கற்றல் மாணவர்களுக்குச் சுவையான அனுபவமாக அமையும். மேலும், சரியான முறையில் இதைக் கையாண்டால் மாணவர்களின் தமிழ்மொழியாற்றலை வளர்க்கலாம்.

சிறப்பு நோக்கம்

கணினிவழிக் கற்றலில் ஈடுபடுவதற்கு முன் மாணவர்களிடம் இருந்த தமிழ்மொழித்திறனுக்கும் கணினிவழிக் கற்றலில் ஈடுபட்டபின்பு காணப்பட்ட தமிழ்மொழித்திறனுக்கும் உள்ள வேறுபாட்டைக் கண்டறிவதே இக்கட்டுரையின் சிறப்பு நோக்கமாகும்.

கருதுகோள்

பொதுநோக்கம் சிறப்பு நோக்கங்களுக்கு ஏற்ப நடவடிக்கைகள் வடிவமைக்கப்பட்டன. கீழ்க்காணும் கருதுகோள் உருவாக்கப்பட்டது.

கணினிவழிக் கற்றல் மாணவரின் நிலைக்கு ஏற்ப தமிழ்மொழியைப் பயன்படுத்தும் ஆற்றலை வளர்க்க உதவுகின்றது.

(கணினிவழிக் கற்றலில் ஈடுபடுவதற்கு முன் இருந்த மொழித்திறனைவிட கணினிவழிக் கற்றலுக்குப் பின் தமிழ்மொழித்திறன் மேம்பட்டுள்ளது.)

ஆய்வினைச் செய்த முறை

புள்ளிவிவரங்களை அடிப்படையாகக்கொண்டு (statistical experiment) இந்த ஆய்வு அமையவில்லை. இவ்வாய்வில் புத்தாக்க முறையில் பயிற்சிகளும் அணுகுமுறைகளும் கையாளப்பட்டன.

ஆய்வு மேற்கொள்ளும்போது கடைப்பிடிக்கப்பட்ட விதிமுறைகள்.

1. ஒவ்வொரு மாணவருக்கும் ஒரு கணினி கொடுக்கப்பட்டது.
2. கணினியை இயக்கப் போதுமான பயிற்சிகள் வழங்கப்பட்டன.
3. பாடப்பொருள் கவனமாகத் தேர்ந்தெடுக்கப்பட்டது.
4. பயிற்றுவனங்கள் தயாரிக்கப்பட்டன.
5. ஆங்கிலப் பாடங்களுக்குத் தேர்ந்தெடுக்கப்பட்ட கருப்பொருளே தமிழ்ப் பாடங்களை உருவாக்கப்பயன்பட்டது.
6. மாணவர்கள் ஒவ்வொருவரும் ஒரு 'thumb-drive' வாங்கிப் பயன்படுத்தினர்.

முன் தயாரிப்பு

தொடக்கநிலை ஒன்றிலிருந்தே மாணவர்களுக்குக் கணினியில் PowerPoint மென்பொருள், PhotoStory3 For Windows மென்பொருள் முதலிய மென்பொருள்களை இயக்கக் கற்றுக்கொடுக்கப்பட்டுள்ளது. மாணவர்கள், PowerPoint மென்பொருளின் ஒலிப்பதிவு கருவியை இயக்கி ஒலிப்பதிவு செய்யவும், ஒலிப்பதிவை ஆசிரியர் கோப்புக்குள் (Save it into the Teachers Shared drive) ஏற்றம் செய்யவும், ஆசிரியர் கோப்பிலிருந்து இறக்கம் (Retrieve the file from the Shared drive) செய்யவும் தங்கள் வேலையைச்

சேமிக்கவும் (Save into Thumbdrive) விரைவில் கற்றுக் கொண்டனர். PhotoStory3 For Windows பயன்படுத்தி மாணவர்கள் புகைப்படங்களைக் கோர்வையாக்கி கதைக்கூறுவர்.

தவணை ஒன்றின் முதல் இரண்டு வாரங்களில் தொடக்கநிலை ஒன்றாம் மாணவர்களின் பேச்சாற்றல் சோதிக்கப்பட்டது. மாணவர்களிடம் சில படங்கள் காட்டப்பட்டன. மாணவர்கள் அப்படங்களில் காண்பனவற்றை விவரித்துச் சொல்ல வேண்டும். மாணவர்களின் சொல்வளம், சரளம், கருத்துரைக்கும் திறன் முதலியன சோதிக்கப்பட்டன. பின் மாணவர்கள் முதல் தவணையில் கற்றல் நடவடிக்கைகளில் ஈடுபட்டனர். அவர்கள் கணினிவட்டுகளில் இடம்பெற்ற கதைகளையும் பாடல்களையும் கேட்டனர். பின் அக்கணினி வட்டுகளில் இடம்பெற்றுள்ள பயிற்சிகளைச் செய்தனர். இரண்டாவது தவணையில் மாணவர்கள் கணினியை இயக்கக்கற்றுக்கொண்டதும் ஒரு மாணவருக்கு ஒரு கணினி (one to one computing) என்ற சூழலில் கற்றல் பயணத்தை மேற்கொண்டனர். படவில்லைகளில் இருந்த படங்களைப் பார்த்துப் பேசினர். பின்னர், அவற்றை ஒலிப்பதிவு செய்தனர். அடுத்து, அவற்றை ஆசிரியர் கோப்புக்குள் மற்றும் தங்களது thumb-driveஇல் ஏற்றம் செய்தனர்.

1. ஆசிரியர் 6 பாடங்களைத் தயாரித்திருந்தார். ஒவ்வொரு பாடத்திலும் ஆசிரியர் புளூம்ஸ் கோட்பாட்டைப் பின்பற்றிக் கேள்விகளை ஒரு படவில்லையிலும் அடுத்த படவில்லையில் அக்கேள்விகளுக்கான பதில்களையும் அமைத்து ஒலிப்பதிவு செய்தார். (கலைமணி 2005, (Chitra Shegar & Ridzuan Bin Abdul Rahim, 2007, Chapter 3)

இப்பயிற்சிகள் மாணவர்களைச் சிந்திக்க வைக்கவும் அவர்களிடம் கேள்விகள் கேட்கும் தன்மையை ஊக்குவிப்பதற்காகவும் தயாரிக்கப்பட்டவை. மேலும் கேள்விகளுக்கான விடைகள் அந்த வாரத்திற்குரிய பாடத்தை ஒட்டி அமைந்த சொற்களாகவும் சொற்றொடர்களாகவும் இருந்தன. அடுத்த படியாக மாணவர்கள் பயிற்சியில் கொடுக்கப்பட்டிருந்த படத்திற்கு ஏற்ப கேள்விகளையும் பதில்களையும் ஒலிப்பதிவு செய்தனர். மாணவர்கள் இப்பயிற்சிகளை முடித்ததும் ஒரு படம் கொண்ட படவில்லையைப் பார்த்துச் சுயமாகக் கேள்வி கேட்டுப் பதில்களை ஒலிப்பதிவு செய்தனர். பின்னர் நான்கு படங்கள் கொண்டமைந்த படத்தொடரைக் கோர்வையாக்கிப் பேசினர்.

ஒரு மாணவரின் ஒலிப்பதிவை சக மாணவர்களும் ஆசிரியரும் கேட்டுத் தங்களது கருத்துகளை ஒலிப்பதிவு செய்தனர். (peer collaboration/ teacher feedback/ peer evaluation) ஆசிரியர் கருத்துரைத்தல், சக மாணவர்கள் மதிப்பிடுதல். பின், அம்மாணவர் தமது படைப்புக்கான பிறரின் கருத்துகளைக் கொண்டும் சுய மதிப்பீட்டைக் கொண்டும் கிடைத்த விமர்சனங்களைக் (feedback / self reflection) கொண்டும் தம்முடைய தவறுகளைத் திருத்திக்கொண்டு மீண்டும் ஒலிப்பதிவு செய்தார். ஆசிரியர் மாணவர்களின் ஒலிப்பதிவுகளைக் கேட்டு மதிப்பீடு செய்தார். (Towndrow, P. A. & Vallance, M. (2004).

அணுகுமுறை

PETALS எனப்படும் ஈடுபாடுமிக்க கற்றல் கோட்பாடுகளைப் பயன்படுத்திப் பாடங்களைத் திட்டமிட்டோம். மாணவர்களின் தயார் நிலை, மாணவர்களின் தேவைகள் மற்றும் கற்றல்பாணி முதலியவற்றை மனத்தில் கொண்டோம் (Allwright, D. & Bailey, K. M. (1991). மாணவர்கள் நடவடிக்கைகளின் வாயிலாகப் பெறக்கூடிய கற்றல் அனுபவங்களை மனத்திற் கொண்டே நடவடிக்கைகள் திட்டமிடப்பட்டுச் செயலாக்கம் கண்டன.

1) நடவடிக்கைகள்

மாணவர்கள் விரைவாகத் தமிழில் சில சொற்களைக் கைவரப் பெற வேண்டும் என்ற எண்ணத்தில் அவர்களுக்குக் கதைத்தொகுப்பு கணினி வட்டுகள் வழங்கப்பட்டன. அவை ஆங்கிலம் தமிழ் ஆகிய இருமொழிகளில் அமைந்த கதைகளைக் கொண்டவை. அவை மாணவர்களின் அனுபவத்திற்கும் அறிவுக்கும் உட்பட்டதாக இருந்தன. இரு வாரங்களுக்கு ஒரு கணினிவட்டை இரவல் பெற்று மாணவர்கள்

தங்கள் வீடுகளில் அவர்களுடைய பெற்றோருடன் அமர்ந்து கணினியில் பார்க்க ஊக்கமளிக்கப்பட்டது. பிறகு, பள்ளியில் IWB (இணைச்செயல் வெண்பலகை) பயன்படுத்தி மாணவர்களுக்குத் தமிழ் எழுத்துக்கள் கற்பிக்கப்பட்டன. பின், தமிழோசை பயிற்றுக்கருவிகளும் பாடமும் பீக்கன் பள்ளி ஆங்கிலத் துறையின் தொகுதிகளுக்கேற்ப தொகுக்கப்பட்டன.

2) "ஒரு மாணவருக்கு ஒரு கணினி" (one to one computing)

தொகுக்கப்பட்ட பாடங்களை மாணவர்கள் ஒரு மாணவருக்கு ஒரு கணினி (one to one computing) என்ற சூழலில் கற்றனர். பல ஒலிப்பதிவுப் பயிற்சிகளின் வழி மாணவர்கள் அடிப்படைச் சொற்களைக் கற்றனர். PowerPoint படவில்லைகளில் 'நேரடியான விடைகளைத் தரக்கூடிய எளிய வினாக்கள் மற்றும் சிந்தனையைத் தூண்டக்கூடிய கடினமான வினாக்கள்' அமைக்கப்பட்டன (Chitra Shegar & Ridzuan Bin Abdul Rahim, 2007, Chapter 3). மாணவர்கள் தயாரிக்கப்பட்ட படவில்லைகள் வாயிலாகக் கற்ற சொற்களைப் பயன்படுத்திப் பேச வாய்ப்புகள் ஏற்படுத்தித்தரப்பட்டன. அவர்கள் தொடர்ந்து பயிற்சிகளை வீடுகளில் செய்வதற்கும் அனுமதி அளிக்கப்பட்டது. ஒரு மாணவருக்கு ஒரு கணினி என்ற சூழலில் மாணவர்கள் சுயமாகத் தங்கள் கற்றல் பயணத்தை மேற்கொண்டனர். மாணவர்கள் ஒரு படத்தைப் பற்றிப் பேசியதும் தொடர்ந்து நான்கு படங்கள் கொண்ட கதை தொடரை வரிசைப்படுத்திப் பேசினர். இந்நடவடிக்கைகளின்வழி மாணவர்கள் புளும்ஸ் கோட்பாட்டின் அடிப்படைநிலைகளான நினைவுகூர்தல் (remembering) புரிந்துணர்தல், (understanding) பொருத்திப்பார்த்தல் (applying) ஆகிய கற்றல் நிலைகளை அடைந்தனர்.

3) அனுபவம் வழிக் கற்றல் (Experiential learning)

தொடக்கநிலை ஒன்றாம் மாணவர்கள், சிங்கப்பூர் விலங்கியல் தோட்டத்திற்குக் கற்றல் பயணம் மேற்கொண்டனர். மாணவர்கள் மிருகங்களின் முகங்களை முகமூடிகளாகச் செய்தனர். இரு குழுக்களும் நாடகம் தயாரித்தனர். ஒரு குழுவினர் அன்று காலையில் பள்ளித் தலைமையாசிரியர் பகிர்ந்து கொண்ட _சுத்தம் சுகம் தரும்_ என்ற கருத்தை மையமாகக் கொண்டு அதுவரை ஒரு மாணவருக்கு ஒரு கணினி (one to one computing) என்ற சூழலில் அவர்கள் கற்ற சொற்களைப் பயன்படுத்திக் கதையை உருவாக்கினார்கள். பின்னர், அதற்கேற்ப வசனங்களை உருவாக்கிப் பேசி நடித்தனர். மற்றொரு குழுவினர், தமிழாசிரியர் பகிர்ந்து கொண்ட _சிங்கமும் முயலும்_ என்ற பஞ்சதந்திரக் கதையை மையமாகக் கொண்டு பொருத்தமான வசனங்களை அமைத்துப் பேசி நடித்தனர்.

மாணவர்கள் 2-ஆம் தவணையின் எட்டாவது வாரத்தில் மற்றொரு நாடகத்தை நடித்துக் காட்டினர். இந்த நாடகத்தின் சாரம் அறநெறிக் கல்வி பண்பான _மரியாதை_ என்னும் பண்பை வலியுறுத்துவதாக அமைந்தது.

4) பெற்றோர் பங்களிப்பு

பெற்றோரின் பங்கு இல்லையேல் எடுத்துக்கொண்ட பணியில் வெற்றிகாண்பது சவாலாக அமையும் என்பதை உணர்ந்து மாணவர்களின் கல்விப்பயணத்தில் அவர்களையும் ஈடுபடுத்தினோம். பெரும்பாலான மாணவர்களின் வீடுகளில் தமிழ்ப் புழக்கம் இல்லை என்பதை உணர்ந்ததால் பல நடவடிக்கைகள் அவர்கள் தங்கள் பெற்றோருடன் இணைந்து ஈடுபடுமாறு அமைக்கப்பட்டன. பெற்றோர்களிடம் வகுப்பறையில் நடத்தப்படும் பாடங்களைப் பற்றியும் நடவடிக்கைகளைப் பற்றியும் தெளிவாகக் கடிதம்வழியும் நேரடிச் சந்திப்பின்போதும் எடுத்துரைக்கப்பட்டது. நடவடிக்கைகளுக்கு இடையே பெற்றோர்கள் சந்தித்த சவால்களைக் குறிப்பெடுத்துக் கொண்டோம். பிறகு தொடர்நடவடிக்கைகளை மறுபரிசீலனை செய்து மாற்றி அமைத்தோம். பெற்றோருடன் இணைந்து ஈடுபடும் வகையில் விடுமுறை வீட்டுப்பாடங்களையும் அமைத்தோம்.

முடிவுகள்

மாணவர்கள், குறிப்பாக ஆங்கிலம் பேசும் பின்புலத்திலிருந்து வரும் மாணவர்கள் இப்பொழுது ஆங்கிலத்தையும் தமிழையும் கலந்து பேச முயற்சி எடுத்துள்ளனர். அவர்கள் கற்றுக்கொண்ட சொற்களைப் பயன்படுத்தி வாக்கியங்களில் கருத்துகளைப் பகிர்ந்துகொள்ள முயன்றுள்ளனர். மேலும் தமிழில் கூறப்பட்ட கதையைக் கேட்டுப் புரிந்துகொண்டு அதன் முடிவை ஊகித்துக்கூறவும் முயன்றுள்ளனர் (Jonassen, D., Howland J., Marra, R.M. & David., 2008). ஓரளவுக்குச் சிறப்பாகப் பேசக்கூடிய மாணவர்களால் தங்கள் சக மாணவர்களின் கருத்துகளைக் கேட்டு மதிப்பீடு செய்ய முடிகிறது. சிறப்பாகப் பேசக்கூடிய மாணவர்களால் பேச்சுத்தமிழைப் பயன்படுத்திப் படங்களைக் கோர்வைப்படுத்திப் பேச முடிந்தது. மிகவும் சிறப்பாக ஆனால் எழுத்துத் தமிழில் பேசக்கூடிய மாணவரால் மிகவும் சிறப்பாகக் கருத்துரைக்கவும் மேலும் பேச்சுத்தமிழில் உரையாடல் நிகழ்த்தவும் முடிந்தது.

சந்தித்த சவால்கள்

1. மாணவர்கள் மதிப்பீடு தரவரிசைப் பற்றியும், எப்படி மதிப்பீடு செய்வது என்பதைப் புரிந்துகொண்டும் சரிவரச் செயல்படச் சிரமப்பட்டனர்.
2. மாணவர்களின் கற்றல் பாணிக்கேற்ப நடவடிக்கைகளை அமைப்பதில் சிரமங்கள் ஏற்பட்டன.
3. பலதரப்பட்ட கணினி இயக்கும் திறன் உள்ள மாணவர்களின் ஆற்றலுக்கேற்ப நடவடிக்கைகளை அமைப்பதில் சிரமம் ஏற்பட்டது.

கணினிவழிக் கற்றலின் பலன்கள்

1. கணினிவழிக் கற்றலில் மாணவர்கள் தாங்கள் ஒலிப்பதிவு செய்ததை மீண்டும் மீண்டும் கேட்டுத் தங்களுடைய தவறுகளைத் தாங்களே திருத்திக் கொள்ள இயன்றது.
2. மாணவர்கள் தங்கள் கற்றலுக்குத் தாங்களே பொறுப்பேற்க (Independent learners) கணினிவழிக் கற்றல் வழிவகுத்துள்ளது (Jonassen, D., Howland J., Marra, R.M. & David, C., 2008).
3. மாணவர்கள் தங்கள் படைப்புகளைச் செய்வதில் பெரும் உற்சாகம் காட்டுகின்றனர். தமிழ் மொழியில் பேசி தங்களது படைப்புகளை இன்னும் சிறப்பாக அமைக்க ஆர்வம் காட்டுகின்றனர். அவர்களுடைய தன்னம்பிக்கையை வளர்க்க உதவுகிறது.
4. ஆசிரியரின் இடையீட்டு நடவடிக்கைகள் தயாரிப்பதற்கு உதவியாக இருக்கிறது.
5. நவீன தகவல் தொழில்நுட்பத்தை (modern technology) பயன்படுத்தும்போது மாணவர்களின் உற்சாக நிலை மேம்படுகிறது.

முடிவுரை

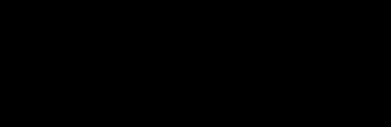
கணினி சுதேசிகள் என்று கூறப்படும் இக்காலத்து மாணவர்களுக்குப் பேசத் தேவையான அடிப்படை மொழிவளமும் சிந்தனையாற்றலை மேம்படுத்த உதவும் கற்றல், கற்பித்தல் நடவடிக்கைகளும் ஒரு மாணவருக்கு ஒரு கணினி_ (one to one computing) என்ற சூழலுக்கேற்ப அமைக்கப்பட்டன. தமிழில் ஒரு சொல்கூடத் தெரியாமல் ஆங்கிலத்திலேயே உரையாடும் தமிழ் மாணவர்களிடையேயும் தமிழ்ப்பழக்கம் குறைந்து காணப்படும் இல்லங்களிலும் தமிழோசை ஒலிக்கக் கணினிவழி கற்றல் வழிவகுத்துள்ளது. மாணவர்கள் வீட்டில் செய்த பயிற்சிகளில் பெற்றோர் ஒத்துழைப்பு இருந்திருக்கலாம். அதுவே அவர்கள் பள்ளியிலும் சிறப்பாகச் செய்ய ஒரு காரணமாக அமைந்திருக்கலாம். இருப்பினும் மாணவரின் திறனைக் கணினிக்கும்போது சோதனைக்குரிய பயிற்சிகள் யாவும் பள்ளியிலேயே நடத்தப்பட்டன. மேலும், கணினிவழிக் கற்றல் வீட்டில் மாணவர்களையும் பெற்றோரையும் இணைக்கும் பாலமாக அமைந்துள்ளது. தமிழ் மொழிப்புழக்கம் பீக்கன் தொடக்கப்பள்ளித் தமிழ் மாணவர்களின் இல்லத்தில் மேம்பட்டுள்ளது என்பது வெள்ளிடைமலை.

நன்றி

ஆய்வுக்கட்டுரை திருந்திய வடிவம் பெறுவதற்கு உறுதுணை புரிந்த பீக்கன் தொடக்கப்பள்ளியின் துணைத் தலைமையாசிரியர் திரு சூராஜ் நாயர் அவர்களுக்கும் குவோ சுவான் பிரஸ்பிடேரியன் உயர்நிலைப்பள்ளியின் மூத்தத் தமிழாசிரியர் திரு கலைமணி அவர்களுக்கும் எங்கள் நன்றி உரித்தாகுக.

References

1. Jonassen. D., Howland J., Marra. R.M. & David. C. Meaningful Learning with Technology (2008). Columbus, Ohio: Merrill Prentice Hall
2. Report of the Chinese Language Curriculum and Pedagogy Review committee (2004). Retrieved May 28, 2008, from <http://www.moe.gov.sg/media/press/2004/>
3. Tan, Y.H., Ow, J & Tan, S.C. (2006). Audioblogging: Supporting the Learning of Oral Communication Skills in Chinese Language. Paper presented at the AECT Research Symposia, Indiana, United States. (pdf, 150KB)
4. Allwright, D. & Bailey, K. M. (1991). Focus on the language classroom: An introduction to classroom research for language teachers. New York: Cambridge University Press.
5. Towndrow, P. A. & Vallance, M. (2004). Using It in the Language Classroom: A Guide for Teachers and Students in Asia. Singapore: Longman Pearson Education.
6. Redesigning Pedagogy – Voices of Practitioners by Chitra Shegar & Ridzuan Bin Abdul Rahim, 2007 . Chapter 3 ‘ Using IT to improve the oral and aural performance of Tamil language students by developing metacognitive skills’, Mr Kalaimani and Mrs Sivagouri Kalamoorthy.



2

இணையவழிக் கல்வி



தமிழ் - கற்றல் கற்பித்தலில் கூகல்

G. Murugan, Subject Head
Bendemeer Secondary School
g_murugan@moe.edu.sg

M. Gnanasekaran, Sr Teacher
CHIJ St. Theresa's Convent
m_gnanasekaran@moe.edu.sg

S. Mohan, Senior Teacher,
Crescent Girls Secondary School,
sambandam_mohan@moe.edu.sg

தகவல் தொழில்நுட்பப் புரட்சியால் நாளுக்கு நாள் தோற்றம் கண்டுவரும் நவீன மென்பொருள்களும், கருவிகளும் எங்கும் எதிலும் துரித மாற்றங்களை ஏற்படுத்தி உலகைத் தன்வயப்படுத்தி வருகின்றன. இப்புரட்சியில் விளைந்த கணினி, இணையம் மற்றும் கைத்தொலைபேசி மனிதனின் வாழ்க்கை முறையோடு ஒட்டி உறவாடி இரண்டறக் கலந்துவிட்டன. மணலில் விரல் தேய “அ” எழுதிக் கற்ற முறையிலிருந்து இன்று Tablet PC யிலும், கருத்துப்பரிமாற்ற வெண்பலகையிலும் (Interactive White Board) விரலால் எழுதிப் பழகும் நிலைக்குக் கற்றலிலும் கற்பித்தலிலும் தகவல் தொழில்நுட்ப வசதிகளைப் பயன்படுத்தத் தொடங்கிவிட்டோம். இதில் தமிழ் கற்றலும் கற்பித்தலும் விதிவிலக்கல்ல.

அவனன்றி ஓர் அணுவும் அசையாது என்பது போல கணினியின்றி ஒரு நிமிடமும் இன்றைய இளைய சமுதாயத்தினரால் வாழ முடியாது என்றநிலை உருவாகி விட்டது. இன்றைய மின்னியல் உலகைப் பொறுத்தவரை மனித சமுதாயத்தை மூன்று தலைமுறைகளாகப் பிரிக்கின்றனர். முதல் தலைமுறை இணையத்தோடு பிறந்து இணையச் சூழலில் வளரும் டிஜிட்டல் தலைமுறை. இரண்டாவது தலைமுறை இணையத்தின் வளர்ச்சியைப் பார்த்து அதனோடு பரிட்சயம் செய்துகொண்ட தலைமுறை. மூன்றாவது தலைமுறை இணையம் என்றாலே பயந்து ஒதுங்கிக்கொள்ளும் மூத்தத் தலைமுறை.4 இன்றைய மாணவர்கள் தகவல் தொழில்நுட்பங்களின் வளர்ச்சியில் பிறந்து வளரும் முதல் தலைமுறையைச் சேர்ந்தவர்கள். இவர்கள் எதனையும் உடனுக்குடன் எதிர்பார்க்கும் சூழலில் வளர்கின்றனர். எனவே இவர்களை ‘டிஜிட்டல் கிட்ஸ்’ என்று அழைக்கின்றனர். இவர்கள் விரும்பும் வகையில் தமிழ் கற்பித்தலைச் சிங்கப்பூர் ஆசிரியர்கள் சிரத்தையோடு கையாண்டு வருகின்றனர்.

கற்றல் கற்பித்தலில் தகவல் தொழில் நுட்பம்

21-ஆம் நூற்றாண்டில் அறிவுசார் பொருளியலுக்கு மாணவர்களைத் தயார்ப்படுத்தும் வகையில் சிங்கப்பூர்க் கல்வியமைச்சு Master Plan 3 ஐ அறிவித்தது.5 அவற்றில் முக்கியமாக நான்கு நோக்கங்கள் வலியுறுத்தப்பட்டுள்ளன. அவைகளாவன:

- சுயக் கற்றல் திறனை வளப்படுத்துல்
- மாணவர்களின் கற்றல் அடைவுகளை அடையாளங்கண்டு அதற்குத் தக்கவாறு கற்பித்தல் உத்திகளையும் வளங்களையும் பயன்படுத்துதல்
- மாணவர்களை ஆழமாகவும், அகலமாகவும் கற்க ஊக்குவித்தல்
- எங்கும் எப்பொழுதும் கற்றல்

4 <http://ezilnila.com/archives/1490>

5 <http://www.moe.gov.sg/media/press/2008/08/moe-launches-third-masterplan.php>

இதனடிப்படையில் ஆசிரியர்கள் பாடப்பொருளையும், கற்பித்தல் முறைகளையும், மதிப்பிடுதலையும் இணைத்துக் கற்பிக்கும் முயற்சியில் ஈடுபட்டுள்ளனர். இதற்கு ஆசிரியரை மையப்படுத்தியக் கற்பித்தல் முறையோடு, மாணவரை மையப்படுத்தியக் கற்பித்தல் (Student Centredness) முறைகள் விரிவாக நடைமுறைப்படுத்தப்பட்டு வருகின்றன.

ஒவ்வொரு பள்ளியும் தகவல் தொழில்நுட்பத்தைப் பல்வேறு வகைகளில் கற்றல் கற்பித்தலுக்குப் பயன்படுத்தி வருகின்றது. கற்பித்தல் நேரத்தில் 30% பாட நேரங்கள் தகவல் தொழில்நுட்பத்தின் அடிப்படையில் அமைந்திருக்க வேண்டும். இதற்குத் தேவையான பயிற்று வளங்களைக் கல்வி அமைச்சின் பாடநூலாக்கக் குழுவும், கல்வித் தொழில்நுட்ப மேம்பாட்டு மன்றமும் (ETD) வழங்கிவருகின்றன. ஓரிரு வணிக நிறுவனங்களும் பயிற்று வளங்களையும் கற்றல் வளங்களையும் இணையத்தின் மூலம் வழங்கிவருகின்றன. ஆனால், ஆசிரியர்களாகிய நாங்களும் எங்களின் மாணவர்களின் கற்றல் திறனுக்கேற்ப, நாங்கள் விரும்பும் வகையில் எப்படித் தரமான கற்றல் கற்பித்தல் வளங்களைத் தயாரிக்கலாம் என அலசியபோது கூகல் ஆவணங்கள் தென்பட்டன. இதை ஏன் நாம் தமிழ் கற்றல் கற்பித்தலுக்குப் பயன்படுத்தக் கூடாது என முயன்று பார்த்தோம். அவற்றைப் பற்றிப் பகிர்ந்துகொள்வதே இக்கட்டுரையின் நோக்கமாகும்.

நோக்கங்கள்

- நவீன தொழில்நுட்ப வசதிகளைப் பயன்படுத்தித் தமிழ் கற்றல் கற்பித்தலில் ஆர்வத்தைத் தூண்டுதல்.
- மாணவர்களுக்கு அளவான கற்பித்தலில் நிறைவான, விரிவான கற்றல் வாய்ப்புகளை உருவாக்குதல்.
- இணையத்தின் வாயிலாக வகுப்பறைக்கு வெளியேயும் சுயக்கற்றலைத் தொடர்வதற்கான வாய்ப்பினை ஏற்படுத்துதல்.
- மாணவர்களின் சிந்தனை ஆற்றலுக்கும் படைப்பாக்கத் திறனுக்கும் உரமூட்டுதல்.
- ஆசிரியர்கள் ஒருங்கிணைந்த முறையில் தேவைக்கேற்ற, காலத்திற்கேற்ற கற்றல் கற்பித்தல் வளங்களை உருவாக்குதல்.

கற்பித்தல் கருவிகளும் பயன்பாடும்

இங்கு நாங்கள் கற்பித்தலுக்குக் கூகல் தளத்தில் உள்ள அஞ்சல் (Mail), உரையாடல் (Talk), ஆவணங்கள் (Docs), தளங்கள் (Sites), பூமி (Earth), மற்றும் பிகாசோ கருவிகளைப் பயன்படுத்திக் கொண்டோம். இவை எல்லாம் ஒரே தளத்தில், அதுவும் இலவசமாகக் கிடைப்பது எங்களுக்கு மிகவும் வசதியாக இருந்தது.

கூகல் அஞ்சல் மூலம் மாணவர்கள் அனைவருக்கும் முகவரி உருவாக்கப்பட்டு, ஒவ்வொரு பள்ளியிலும் உள்ள ஒரே வகுப்பைச் சேர்ந்த மாணவர்கள் குழுக்களாக்கப்பட்டனர். இது இருவழித் தகவல் பரிமாற்றங்களுக்கு மிகவும் உதவியாக இருந்தது. ஆசிரியர் தர விரும்பும் பயிற்சிகளையும், கற்றல் வளங்களையும் மாணவர்களுக்குத் தெரியப்படுத்தவும், மாணவர்களின் ஐயங்களைப் போக்கவும் முடிந்தது. கூகல் உரையாடல் மாணவர்கள் பாடங்களைச் செய்யும்போது அவர்களுக்கு ஏற்படும் ஐயப்பாடுகளைப் போக்கிக் கொள்ளவும், தமிழைப் பயன்படுத்தும்போது ஏற்படும் தொழில்நுட்பச் சிக்கல்களைப் போக்கிக்கொள்ளவும் பயன்படுத்தப்பட்டது. ஆவணங்கள் பகுதியில் ஆசிரியர்கள் மாணவர்களுக்குத் தர வேண்டிய பயிற்சிகளையும், பாடக் குறிப்புகளையும், வகுப்பில் நடைபெற்ற பாடங்கள் தொடர்பான தகவல்களையும் ஏற்றம் செய்தோம். அவற்றை மற்ற பள்ளி மாணவர்களும் ஆசிரியர்களும் பயன்படுத்திக்கொள்ளச் செய்தோம். ஆவணத்திலுள்ள கருத்தளிப்புப் பகுதியை (Forum)

மாணவர்களுக்கான பயிற்சியைத் தயார் செய்து அனுப்புவதற்கும், மாணவர்கள் அனுப்பும் விடைகளைச் சேகரிக்கவும் பயன்படுத்திக்கொண்டோம்.

கூகல் தளத்தில் எங்கள் இணையப் பூங்கா உருவானது. பூஞ்சோலை (<http://sites.google.com/site/malargal3/>) என்ற பெயரில் பொதுவான இணையப் பக்கத்தையும், அதன் இணைப்பாக நாங்கள் மூவரும் தாமரை, மல்லிகை, முல்லைத் தோட்டங்களையும் (இணையப் பக்கங்களை) உருவாக்கியுள்ளோம். அவை தற்போது அரும்புவிடத் தொடங்கியுள்ளன. அவை இன்னும் பலவாகப் பூத்து மணம் வீசும்பொழுது பூங்காவுக்கு வருகின்றவர்களை மணம் மகிழ்ச்சியும் என்பது உறுதி. வடிவமைப்பும் வளங்களும் அவரவர்களின் விருப்பத்திற்கேற்ப தனித்தனியே இருந்தாலும் பயன்பாடு ஒருங்கிணைக்கப்பட்டது. மொழிக் கற்றலில் அடிப்படைத் திறன்களான நோக்கல், கேட்டல், பேசுதல், வாசித்தல், எழுதுதல், ஆகியவற்றில் மாணவர்கள் தங்கள் திறனை வளர்த்துக் கொள்ளும் வகையில் வளங்கள் சேர்க்கப்பட்டுள்ளன. ஆசிரியர்கள் தயாரித்த மற்ற வலையொலி (podcast), வலைப்பூக்கள் (blogs), படத்தொகுப்புகள் பிகாசோ, யூடியூப் ஒளிப்படங்கள் முதலானவற்றைத் தனித்தனியே தயாரித்திருந்தாலும் ஒன்றிணைக்க முடிகிறது.

மாணவர்களின் செயல்திறனை மேம்படுத்தல்

மூன்று ஆசிரியர்கள், மூன்று பள்ளிகள் இதில் இணைந்திருப்பதால் அவர்களுக்கிடையே உள்ள செயல்திறனை ஒப்பிட்டுப் பார்க்க முடிகிறது. ஒருங்கிணைந்த கற்றல் என்ற நிலை இங்கே செயல்படுத்தப்படுகிறது. ஒரு பள்ளி ஒரு வகுப்பு என்ற நிலையிலிருந்து வேறுபட்டு, மேம்பட்டு வெவ்வேறு சூழல்களில் வெவ்வேறு பள்ளிகளில் உள்ள மாணவர்கள் தாங்கள் செய்த பயிற்சிகளை, விடைகளின் தரத்தை, சிறப்பை மற்ற பள்ளி மாணவர்களுடன் ஒப்பிட்டுப் பார்த்து மதிப்பிட முடிகிறது. அவ்வாறு மதிப்பிடுவதன் மூலம் தங்களை மேம்படுத்திக் கொள்கின்றனர்.

சிங்கப்பூர்க் கல்வித்திட்டத்தில் கேட்டல் திறனுக்குக் கூடுதல் கவனம் செலுத்தப்படுகிறது. மாணவர்கள் தங்களின் வாசிப்புத் திறனையும் கேட்டல் திறனையும் மேம்படுத்திக்கொள்ள வலையொலிப் பக்கங்களை உருவாக்குகிறார்கள். அவற்றை ஒருங்கிணைத்துக் கொடுக்கின்றோம். இது மாணவர்கள் தங்களின் வாசிப்புத் திறனையும், கேட்டல் திறனையும் பிறருடன் ஒப்பிட்டுப் பார்த்து மேம்படுத்திக்கொள்ள வாய்ப்பாக அமைகிறது. இதில் செய்திகள், நாட்டு நடப்புகள், கதைகள், நேர்முக வர்ணனைகள், கலந்துரையாடல்கள் எனப் பல்வேறு அங்கங்கள் இடம்பெறுவதால் சூழலை ஒட்டிய மொழிப்பழக்கத்தையும் வளர்க்க முடிகிறது.

பல்வேறு கற்றல் கற்பித்தல் அணுகுமுறைகளில் கூகல்

கூகல் தளம் ஆசிரியர்களின் பல்வேறு கற்பித்தல் உத்திகளுக்கும், பாடப் பொருள்களுக்கும், மதிப்பீட்டு முறைகளுக்கும் பலதரப்பட்ட வாய்ப்பு வசதிகளை உள்ளடக்கியதாக உள்ளது. வாழ்நாள் கற்றல் என்பதற்கேற்ப மாணவர்களோடு ஆசிரியர்களும் சேர்ந்து கற்பதற்கும் பல்வேறு திறன்களைப் பெருக்கிக் கொள்வதற்கும் ஏராளமான வாய்ப்புகள் உள்ளன. இங்குக் கட்டுரையின் சுருக்கம் கருதி மூன்று கற்பித்தல் உத்திகளை மட்டுமே குறிப்பிட்டுள்ளோம்.

இணைந்து கற்றலும் (Colobarative Learning) படைப்பாக்கத்திறனும்

மாணவர்கள் தங்கள் பள்ளி மாணவர்களுடன் மட்டுமின்றி மற்ற பள்ளி மாணவர்களுடனும் ஒருங்கிணைந்து கற்றல் நடவடிக்கையில் ஈடுபடுவதற்கும், கருத்துப் பரிமாற்றம் செய்வதற்கும், இத்தளம் பெரும் உதவியாக இருக்கிறது. கட்டுரை எழுதும்போது ஒரே மன வரைபடத்தில் மாணவர்கள் கருத்துதிர்ப்பு செய்வதற்கும், ஒருவர் கூறிய கருத்துகளைச் சரியான வகையில் செம்மைப்படுத்துவதற்கும், மாற்றியமைப்பதற்கும், அதனை ஒட்டியோ மறுத்தோ தங்களது கருத்துக்களைக் கூறுவதற்கும் முடிகிறது. இதனால் ஒரு வகுப்பில் உள்ள மாணவர்களோடு இணைந்து கற்ற நிலைமாறி வகுப்பறையிலிருந்துகொண்டே வெவ்வேறு பள்ளி

மாணவர்களுடன் இணைந்து கருத்துதிர்ப்பு செய்து கட்டுரை எழுதும் வாய்ப்பை மாணவர்கள் பெறும்போது கற்றல் ஆழமானதாகவும், சிந்தனையை விரிவுபடுத்துவதாகவும் அமைகிறது.

மாணவர்கள் கட்டுரையை முழுமையாக்கி ஏற்றம் செய்கிறபோது கருத்து ஒன்றாக இருந்தாலும் ஒவ்வொருவரின் மொழித்திறனும், வெளியிடும் முறையும், எவ்வகையில் உயர்ந்திருக்கிறது என்பதைச் சுலபமாக ஒப்பிட்டுப் பார்க்கும் நிலைக்கு உயர்த்தப்படுகின்றனர். இதே போன்று, பிரச்சினைக்குத் தீர்வு காணும் தலைப்புகளில் ஒரு பிரச்சினையைப் பற்றி விவாதிக்கும்போது அங்குப் பல்வேறு தீர்வுகள் பிறக்கின்றன. பள்ளிகளுக்கிடையே ஆரோக்கியமான போட்டித் தன்மையோடு கற்றல் நிகழ்கிறது. இந்த ஒருங்கிணைந்த கற்றலின் வழி பல்வேறு நடவடிக்கைகளை நாம் இத்தளத்தின் வழியாக செயல்படுத்த இயல்கிறது.

வேறுபடுத்தப்பட்ட கற்பித்தல் அணுகுமுறைகள் (Differentiated Instructions)

எல்லா மாணவர்களும் ஒரே மாதிரியாக இருப்பதில்லை. விருப்பத்தில், தயார் நிலையில், கற்கும் வேகத்தில் வேறுபட்டவர்கள். அப்படியிருக்கும் போது கற்பித்தல் முறைகளும் ஒரே மாதிரியாக இருப்பதில் அர்த்தமில்லை. அந்த வகையில் தயாரித்திருக்கும் வளங்களை மாணவர்களின் திறனறிந்து ஆசிரியர்கள் பரிந்துரைக்க முடிகிறது. அதனால் மீத்திறன் மாணவர்களையும், சராசரி மாணவர்களையும், திறன் குறைந்த மாணவர்களையும் ஒரே நேரத்தில் கற்றலில் ஈடுபடுத்த முடிகிறது. இவ்வகையில் இது வேறுபடுத்தப்பட்ட கற்பித்தல் அணுகுமுறைகளைக் கையாள மிகவும் உதவியாக இருக்கிறது.

சுயவிடைக் கருத்தறிதல் பகுதியில் மாணவர்கள் ஒருவர் அனுப்பிய விடைகளை மற்றவர்கள் படித்துப் பார்க்க முடிகிறது, இதனால் ஒரு மாணவன் பகுப்பாய்வு முறையில் ஒரு வினாவிற்கு எத்தனை விதமான விடைகளை அளித்துள்ளனர் என்பதையும், அவற்றில் எது மிகச் சரியானது என்பதையும் பகுத்துணர முடிகிறது. மேலும் இது தவறுகளைச் சுயமாகக் கண்டறிந்து திருத்திக்கொள்ளவும், தன்னைத்தானே சுய மதிப்பீடு செய்துகொள்ளவும் வாய்ப்பாக இருக்கிறது. கட்டுரைகளை அவர்களால் எழுத முடிகிறது. கட்டுரையோடு வேறு சில தகவல்களையும் கொடுத்து விவாதிக்கச் செய்ய முடிகிறது. இதன் மூலம் மாணவர்களின் (Analytical skill) பகுத்தறியும் திறன் வளர்க்கப்படுகிறது

அனுபவக் கல்வி (Experiential Learning)

கூகல் பூமியின் மூலம் மாணவர்கள் நம் கலாச்சாரத் தொடர்புடைய இடங்களையும் வளங்களையும் கற்றுணர முடிகிறது. பாடப்பகுதியில் இடம்பெற்றுள்ள தஞ்சாவூர், மாமல்லபுரம், ஊட்டி, தாஜ்மகால், நயாகரா நீர்வீழ்ச்சி, பர்முடா முக்கோணம், சூயஸ் கால்வாய் போன்ற இடங்களைப் பற்றிய பாடங்களை அனுபவக் கல்வி (Experiential Learning) வாயிலாகக் கற்க முடிகிறது. வெளிநாட்டிற்குச் செல்ல வாய்ப்பில்லாத மாணவர்கள் நேரிடையாக நம்மால் காணமுடியவில்லையே என்ற ஏக்கத்தைப் போக்கிக்கொள்ளவும், துணைக்கோள உதவியுடன் எடுக்கப்பட்ட வரைபடங்களைக் காணவும் முடிகிறது. கூகல் பூமியில் விக்கிபீடியா தகவல்கள், படங்கள், நேரம், வானிலை விவரங்கள், யுடியூப் தகவல்கள், சாலைகள், பாலங்கள், நிலவியல் அமைப்புகள் போன்றவற்றை வகுப்பறையில் இருந்துகொண்டே கண்டு, கேட்டு இரசிக்க முடிகிறது.

இழப்பினைத் தவிர்த்தல்

மாணவர்கள் என்று சொன்னால் எல்லோரும் எல்லா நாளும் தவறாமல் பள்ளிக்கு வருவார்கள் என்றோ வரமுடியும் என்றோ எதிர்பார்க்க முடியாது. அவ்வாறு வராமல் இருக்கும் போது இழப்புகளையும் தவிர்க்க முடியாது. இதனை ஈடுகட்ட என்ன செய்யலாம் என்று யோசித்த போது பாடங்களையும் அதற்குப் பயன்படுத்திய பட வில்லைகளையும் ஆவணங்களையும் அங்கே இடம் பெறச் செய்யலாம் என்று யோசித்தோம்; முடிவெடுத்தோம்; செயல்படுத்தினோம். இதனால் பள்ளிக்கு வராத மாணவர்கள்கூட

தவறவிட்ட பாடங்களை இணையத்தின் இணைப்பு இருந்தால் எங்கும் எப்பொழுதும் பெற முடியும். இதன் மூலம் மாணவர்களின் சுயமாகக் கற்கும் ஆற்றலும் (Self Directed Learning) மேம்படுகிறது.

இன்று ஒவ்வொரு பள்ளியிலும் அப்பள்ளிக்கென்று தனிப்பட்ட இணைய முனையம் (portal) உள்ளது. இருந்தாலும் அவற்றைப் பயன்படுத்த அப்பள்ளி அளிக்கும் மறைச் சொல் இருக்க வேண்டும். மேலும் அப்பள்ளியில் பயிலும் மாணவர்கள் மட்டுமே அதனைப் பயன்படுத்த முடியும். நேரக்கட்டுப்பாடும் இருக்கும். ஆனால் இங்கே மாணவர்களை மையமாகக் கொண்டே இதனை நாங்கள் தயாரித்திருப்பதால் யாரும் எங்கும் எந்நேரத்திலும் கற்றலைத் தொடரலாம்.

வளங்களைப் பகிர்ந்து கொள்ளுதல்

வளங்களைப் பகிர்ந்து கொள்ளுதல் இரு நிலைகளில் சாத்தியமாகிறது. ஒன்று ஆசிரியர்கள் தாங்கள் தயாரித்த பாடங்கள், பாடத்திட்டங்கள், பாடத் தயாரிப்புகள், பாடத்திற்கான துணைக்கருவிகள் போன்றவற்றைப் பிற ஆசிரியர்களுடன் பகிர்ந்துகொள்வதாகும். இரண்டாவது ஒரு ஆசிரியர் தயாரிக்கும் பாடங்கள், பயிற்சித்தாளர்கள், கட்டுரைகள், தேர்வுக்கான குறிப்புகள் போன்றவற்றை மற்றப் பள்ளி மாணவர்களுடன் பகிர்ந்துகொள்வதாகும்.

சிங்கப்பூர்ச் சூழலில் ஆசிரியர்கள் புதியப் புதிய பயிற்றுக் கருவிகளையும், மாணவர்களின் கற்றல் திறனுக்கேற்ற பயிற்று வளங்களையும் தயாரிப்பதற்கு நிறைய நேரங்களைச் செலவிடுகின்றனர். அவை பகிர்ந்துகொள்ளப்படாமல் தனித்தனி நிலையிலேயே இருக்கின்றன. இவற்றிற்கு ஒருநல்லத் தீர்வாக கூகல் ஆவணங்கள் அமைந்துள்ளன. ஆசிரியர்கள் தாங்கள் தயாரித்த பாடங்கள், பாடத்திட்டங்கள், பயிற்று வளங்கள், பயிற்சித்தாளர்கள் போன்றவற்றைக் கூகல் ஆவணத்தில் ஏற்றம் செய்வதன் மூலம் விரும்பிய பகுதியை விரும்புகின்றவர்களுடன் எளிதில் பகிர்ந்துகொள்ள முடிகிறது.

மேலும் ஒரு பாடத்திற்கு ஒரு ஆசிரியர் தயாரித்த அதே வளங்களைப் பிற ஆசிரியர்களும் நேரத்தைச் செலவழித்துத் தயாரிக்க வேண்டியதில்லை. ஒரு ஆசிரியர் தயாரித்த பாடங்களை மற்ற ஆசிரியர் தனது தேவைக்கேற்ப மாற்றங்களைச் செய்து பயன்படுத்திக்கொள்ள முடிகிறது. இதனால் பொன்னான நேரம் மிச்சமாகிறது. அதே சமயத்தில் ஒவ்வொருவரும் பகிர்ந்துகொள்ளும்போது வளங்களும் குவிகின்றன. கற்றல் திறனடிப்படையில் (Learning ability) கற்பிக்க ஏராளமான வளங்கள் கிடைக்கின்றன.

ஆசிரியர் திறன் மேம்பாடு

இன்றைக்கு ஆசிரியர்கள் தங்கள் பணியில் திறனை மேம்படுத்திக் (Professional Development) கொள்ள பல்வேறு பயிற்சி வகுப்புகளுக்குச் செல்கின்றனர். இங்கே இந்தத் தயாரிப்பின் மூலம் ஆசிரியர்கள் தங்கள் பணியை மேம்படுத்திக் கொள்ளவும், ஏராளமாகக் கற்றுக் கொள்ளவும் முடிகிறது. மாணவர்களுக்கான வினாத்தாள்கள், பயிற்சித் தாள்கள் ஆகியவற்றைத் தயாரித்து ஏற்றம் செய்யமுடிகின்றது. மாணவர்கள் அளித்த விடைகள் எல்லாம் ஆன்லைனில் இருப்பதால் அங்கேயே மாணவர்களின் செயல்திறனை மதிப்பிடுவது ஒருபக்கம். இன்னொரு பக்கம் ஆசிரியர்கள் தங்களின் கேள்வித்தாளின் தரத்தை, கடினத் தன்மையை மற்ற பள்ளி மாணவர்களும் விடையளித்திருப்பதிலிருந்து அறிந்து கொள்ள முடிகிறது. அடுத்த கேள்வித்தாளையோ பயிற்சித்தாளையோ தயாரிக்கும் போது வினாக்களின் கடினத் தன்மையை மேம்படுத்தவோ அல்லது மாணவர்களின் நிலைக்கு ஏற்றாற்போல் மாற்றவோ முடிகிறது. இதனால் ஆசிரியர்கள் தங்களின் பணித்திறனை மேம்படுத்திக்கொள்ள முடிகிறது.

ஆசிரியர் கற்பித்தலுக்கான படவில்லைகள் (Power Point) தயாரிக்கும்போது தேவையான யுடியூப் படக்காட்சிகளைப் பயன்படுத்தவும், சுலபமாக இணைப்புக் கொடுக்கவும் பார்க்கவும் முடிகிறது. வழக்கமாக இவ்வாறு சேர்க்கும்போது சேமிப்பதும், பிறரோடு பகிர்ந்துகொள்வதும் சிரமமாகவே இருந்தது. இப்பொழுது அந்தச் சிக்கல்களுக்குத் தீர்வு காணப்பட்டுள்ளது.

இக்காலத்தில் மாணவர்களும் சரி, பெற்றோர்களும் சரி அதிகமாகப் பயிற்சித்தாள்கள் செய்தால்தான் தங்களைத் தேர்வுக்கு உரிய முறையில் நன்றாகத் தயார்ப்படுத்திக் கொள்ள முடியும் என்று எண்ணுகின்றனர். இந்த எண்ணத்தையும் ஈடேற்றக் கூடியதாக கூகல் ஆவணங்களும், கூகல் படிவங்களும் அமைந்துள்ளன.

நிறைவாக...

தமிழ் கற்பித்தலில் தகவல் தொழில்நுட்பத்தைப் பயன்படுத்துவதில் ஆர்வமும் துடிப்பும் மிக்க ஆசிரியர்கள் அவ்வப்போது வருகின்ற தொழில்நுட்பக் கருவிகளைப் பயன்படுத்திப் பல பயிற்று வளங்களைத் தயாரித்து கற்பித்து வருகின்றனர். தொழில்நுட்பங்கள் மாறும்பொழுது தயாரித்த வளங்கள் பயனற்றுப்போகின்றன அல்லது பயன்படுத்த முடியாமல் போகின்றன. இத்தகைய பிரச்சினைகளைக் களைய கூகல் தளம் கைகொடுக்கிறது. ஒரு முறை தயாரித்த வளங்களை எப்பொழுது வேண்டுமானாலும், எங்கு வேண்டுமானாலும் பயன்படுத்திக்கொள்ள ஏதுவாக உள்ளதால் நீண்டகாலத்திற்குப் பயனளிக்கும் என நம்பலாம்.

கற்றல் கற்பித்தல் என்பது வகுப்பறை என்ற எல்லையைக் கடந்த ஒன்றாகிவிட்டது. காலம், நேரம், இடம் என்ற எல்லைகளோ வரையறைகளோ அதற்கில்லை. இன்றைய நிலையில் இணையத்தளங்கள் இருவழிக் கருத்துப் பரிமாற்றம் செய்யும் வகையில் அமைந்துள்ளன. அதனால் மாணவர்களின் உளவியலை அவர்களின் கருத்துகள் வழி அறிந்து அதற்கேற்ப பாடங்களை வடிவமைக்கவும் முடிகிறது. கருத்துப் பரிமாற்றம் ஆர்வம் தரும் வகையில் அமையும் போது கற்றல் வெற்றி பெறுகிறது. ஆசிரியர்கள் மூவர் இணைந்து கூகல் ஆவணத்தைப் பயன்படுத்திய இந்தக் கற்பித்தல் முறையில் பெரும் வளங்களும் பகிர்தலும் இருப்பது அனுபவப்பூர்வமான உண்மை. இதில் மேலும் பல ஆசிரியர்கள் ஒன்றிணைந்து ஈடுபடும்போது அது நாடுகளைக் கடந்து உலகில் வாழும் தமிழ் மாணவர்களுக்கும், ஆசிரியர்களுக்கும், நம் தமிழ் மொழிக்கும் பயன் நல்குவதாக இருக்கும் என்பது வெள்ளிடைமலை.

தமிழ் இணையப் பல்கலைக்கழக மென் பொருள்கள்

ஒரு கண்ணோட்டம்

முனைவர்.ப.அர.நக்கீரன்

இயக்குநர், தமிழ் இணையப் பல்கலைக்கழகம்

1.0 முன்னுரை

இன்று ஆங்கிலத்திலும், பிற மொழிகளிலும் பல மென்பொருள்கள் பயன்பாட்டில் உள்ளன. ஆங்கிலத்திலிருந்து பிறமொழிகளில் மொழிபெயர்க்கும் வசதி, எழுத்தைப் பேச்சாகவும், பேச்சை எழுத்தாகவும் மாற்றும் வசதி, எழுத்துப் பிழைகளைத் திருத்தும் வசதி, அச்சிட்ட பக்கங்களைப் படியெடுத்துப் பதிப்பிக்கும் வகையில் எழுத்துக்களை அடையாளம் காணும் வசதி எனப் பல மென்பொருள்கள் ஆங்கில மொழியில் உண்டு. அவற்றைப் போலவே தமிழிலும் மென்பொருள்களைக் கொண்டு வரும் முயற்சிகள் பல ஆண்டுகளாகத் தொடர்ந்து நடைபெற்று வருகின்றன. இம்முயற்சியில் பல தனியார் நிறுவனங்களும், அரசு நிறுவனங்களும், தமிழ் இணையப் பல்கலைக்கழகமும் ஈடுபட்டுள்ளன.

தமிழக அரசும் தமிழ் மென்பொருள் வளர்ச்சிக்கெனத் தனியே நிதி ஒதுக்கியுள்ளது. அத்திட்டத்தின் கீழ் தமிழ் இணையப் பல்கலைக்கழகம் தமிழ் மென்பொருள்கள் உருவமைக்க நிதி வழங்கி வருகிறது. அதன் பயனாகப் பல மென்பொருள்கள் உருவாக்கப்பட்டுள்ளன. அம்மென்பொருள்கள் பற்றிய ஒரு கண்ணோட்டமே இக்கட்டுரையாகும்.

2.0 மென்பொருள் நிதியம்

தமிழ் இணையப் பல்கலைக்கழகம் தமிழில் மென்பொருள் உருவாக்கத் தமிழ் மென்பொருள் நிதியம் என்ற அமைப்பை ஏற்படுத்தி, செயல்படுத்தி வருவதின் முக்கியக் குறிக்கோள் இணையத் தமிழின் முன்னேற்றத்துக்குப் பாடுபடும் தனி நபர்களையும், நிறுவனங்களையும் கல்லூரிகளையும் கண்டறிந்து அவர்களை ஊக்குவித்து, ஓரளவு நிதியும் வழங்கி மென்பொருள்கள் உருவாக வழி வகுப்பதே ஆகும். மென்பொருள்கள் உருவாக நிதி வழங்கினாலும், அந்த மென்பொருள்களை முழுமையாக உருவாக்கி விற்பனை செய்யும் உரிமை உருவாக்குபவர்களைச் சாரும்.

தமிழகத்தில் உள்ள தனிநபர்களும், நிறுவனங்களும் மைய அரசின் நிதியுடன் செயல்படும் மற்ற மாநிலங்களில் உள்ள நிறுவனங்களும் இத்திட்டத்தில் பங்கேற்கலாம். அண்ணா பல்கலைக்கழகம், அண்ணாமலைப் பல்கலைக்கழகம் மற்றும் அரசு கல்லூரிகள், தனியார் கல்லூரிகள், இந்திய அறிவியல் நிறுவனம், பெங்களூரு, தனியார் மென்பொருள் நிறுவனங்கள் ஆகியவை இத்திட்டத்தில் பங்கேற்று உள்ளன.

தமிழ்நாடு அரசு நடத்திய தமிழ் இணையம் 99 மாநாட்டில் உருவாக்கப்பட்ட குறியீட்டுத் தரப்பாட்டுக்கு ஏற்ப மென்பொருள்கள் உருவாக்கப்பட வேண்டும்.

பேச்சறிதல் (Speech recognition) எழுத்தறிதல் (Optical character recognition), மொழி ஆய்வு (Natural language processing), அகராதிப் பயன்பாடு ஆகிய தனித்த நீண்டநாள் பயன்தரத்தக்க மென்பொருள்களுக்கே முன்னுரிமை அளிக்கப்படுகிறது.

பனுவல் செயலாக்கம் (Text processing), மின்னஞ்சல் பயன்பாடுகள் போன்றவற்றிற்கு நிதி உதவி அளிக்கப்படுவதில்லை.

3.0 செயல்முறை

தனிநபர்களிடமிருந்தும், நிறுவனங்களிடமிருந்தும் திட்டப் பணிகள் வரவேற்கப்படுகின்றன. அளிக்கப்பட்ட திட்டப் பணிகளை ஆராய்ந்து, தேர்ந்தெடுத்து, அறிவுரைகள் கூறி, திருத்தங்கள் செய்து, செம்மைப்படுத்தி, உரிய நிதியைப் பரிந்துரை செய்ய தமிழ் மென்பொருள் வல்லுநர் குழு ஒன்று அமைக்கப்பட்டுள்ளது. சிறந்த கணிப்பொறி மென்பொருள் வல்லுநர்கள் பலர் இதில் உறுப்பினர்களாக உள்ளனர்.

ஒரு திட்டப்பணிக்கு அதன் தகுதிக்கேற்ப உரூ.50,000/- முதல் உரூ.10,00,000/- வரை வழங்கப்படுகிறது. ஒவ்வொரு திட்டப் பணியும் பல கட்டங்களாகப் பிரிக்கப்பட்டு, தொடக்க நிதியாக மொத்தம் வழங்கப்பட உள்ள நிதியில் 20 முதல் 30 விழுக்காடும், பின்னர் ஒவ்வொரு கட்டம் நிறைவடைந்ததும், குழுவின் பரிந்துரைகளுக்கேற்ப நிதி பிரித்து வழங்கப்படும். ஆறு மாதங்கள் முதல் 2 ஆண்டுகள் வரை, திட்டத்தின் தேவைக்கேற்பக் காலம் ஒதுக்கப்படுகிறது. திட்டம் முடிவடைந்ததும், முழுமையாக அது சோதனை செய்து சரிபார்க்கப்பட்ட பின்னரே இறுதித் தவணை நிதி வழங்கப்படுகிறது.

4.0 தமிழ் மென்பொருள்கள்

இதுவரை தமிழ் மென்பொருள் நிதி உதவியால் உருவாக்கப்பட்ட / உருவாக்கப்படும் மென்பொருள்களின் பட்டியல் கீழே கொடுக்கப்பட்டுள்ளது.

4.1 தமிழ் மென்பொருள் வளர்ச்சி நிதியின் கீழ் உள்ள திட்டங்கள்

முடிந்த திட்டங்கள்

திட்டத்தின் பெயர்/ நிறுவனத்தின் பெயர்	வழங்கிய காலம்	அளிக்கப் பட்ட காலம்	முடிந்த காலம்	அளிக்கப் பட்ட நிதி
இதம் 2000 - விண்டோஸ் 2000 மேம்படுத்தப்பட்ட தமிழ் இடைமுகம்	நவம்பர் 1999	6 மாதங்கள்	நவம்பர் 2005	10.00 இலட்சம்
பொன்விழி - தமிழ் எழுத்துரு அறிவி (Tamil OCR)	டிசம்பர் 1999	1 வருடம்	ஜூலை 2002	3.5 இலட்சம்
A PC based Speech Synthesis in Tamil - சொல்லை ஒலிவடிவமாக மாற்றுவதல்	டிசம்பர் 1999	1 வருடம்	அக்டோபர் 2004	4.5 இலட்சம்
தமிழில் விண்டோஸ் 95 - 98	ஜனவரி 2000	6 மாதங்கள்	நவம்பர் 2005	5.0 இலட்சம்
“தமிழ்ப் பொறி” ஆங்கிலம் தமிழ் மொழிபெயர்ப்பு மென்பொருள் (Machine Aided English to Tamil Translation)	மே 2002	6 மாதங்கள்	அக்டோபர் 2008	4.0 இலட்சம்
தமிழ் சொல்வலை	பிப்ரவரி 2003	1 வருடம்	ஆகஸ்ட் 2006	4.0 இலட்சம்
‘ழ’ கனிணி - தமிழில் லைனக்ஸ்	மே 2003	6 மாதங்கள்	நவம்பர் 2004	0.75 இலட்சம்
தமிழ் கைப்பிரதி எழுத்துரு அறிவி	ஜூன் 2003	1 வருடம்	நவம்பர் 2006	5.0 இலட்சம்
செல்லிடப் பேசியில் தமிழ் குறுஞ்செய்தி	பிப்ரவரி 2005	1 வருடம்	ஜூலை 2006	2.0 இலட்சம்
கையடக்கச் சாதனங்களில் கூடுதல் மதிப்பு சேவை - தமிழ் மொழியில்	பிப்ரவரி 2005	1 வருடம்	செப்டம்பர் 2007	3.2 இலட்சம்

உருவாக்கத்தில் உள்ள திட்டங்கள்

திட்டத்தின் பெயர்/ நிறுவனத்தின் பெயர்	வழங்கிய காலம்	அளிக்கப் பட்ட காலம்	முடிந்த காலம்	அளிக்கப் பட்ட நிதி
ஜாவா மூலம் எழுத்துருவிலிருந்து பேச்சு உருவாக்கி (Java-based Tamil Text- To-Speech Synthesizer)	ஜனவரி 2000	6 மாதங்கள்	சனவரி 2010	4.5 இலட்சம்
தமிழ்த் தரவு மேலாண்மைப் பொதி (Tamil Database Management Package)	பிப்ரவரி 2003	30 மாதங்கள்	சனவரி 2010	2.5 இலட்சம்
தகவல் தள மின் அகராதி (தமிழ் - ஆங்கிலம்)	பிப்ரவரி 2005	2 வருடம்	2010	3.465 இலட்சம்
தமிழ் பேச்சு மொழியிலிருந்து எழுத்துரு உருவாக்கி Speech to text	பிப்ரவரி 2005	18 மாதங்கள்	சனவரி 2010	1.40 இலட்சம்
Tamil Corpus Analysis Tools சொல் ஆய்வுக் கருவிகள்	ஜூலை 2006	18 மாதங்கள்	மார்ச் 2010	3.75 இலட்சம்
செல்லிடப் பேசியில், தமிழில் முன்கூறு எழுத்தறிவி மென்பொருள் (Tamil prediction text)	ஜூலை 2006	16 மாதங்கள்	நடைபெற்று வருகிறது	0.711 இலட்சம்
செல்லிடப் பேசியில் நுண்ணறிவுத் தகவல் பெறும் அமைப்பியல் - திருக்குறள் மட்டும் (Thirukkural in mobile phones)	ஜூலை 2006	12 மாதங்கள்	மார்ச் 2010	3.519 இலட்சம்

4.2 தமிழ் எழுத்துக்களைப் படித்துப் பதிப்பித்தல் (OCR for Tamil Text)

அச்சிட்ட தமிழ்ப் பக்கங்களைப் பார்த்து, எழுத்துக்களை அடையாளம் கண்டு, மீண்டும் அப்பக்கங்களை திருத்தி பதிப்பிக்கும் வகையில் புதிதாகத் தட்டச்சுச் செய்ததைப் போலத் தருதல் (OCR for printed Tamil Text). இதனால் மீண்டும் தட்டச்சுச் செய்து பிழைதிருத்தும் பணி தேவைப்படாது.

4.3 ஆங்கிலம் - தமிழ் மொழி பெயர்ப்பு (English - Tamil Translation)

இதுவரை இதனால் சொற்களையும், சிறு சிறு தொடர்களையும் மொழிபெயர்க்க முடிகிறது. நீண்ட தொடர்களை மொழிபெயர்க்க முயற்சிகள் மேற்கொள்ளப்பட்டு வருகின்றன.

30,000 ஆங்கிலச் சொற்களுக்கு ஈடான தமிழ்ச் சொற்களும், 20,000 இந்திய நகரங்களின் பெயர்களும், 1000 சொல் தொடர்களும், பழமொழிகளும் கொண்டு இம்மென்பொருள் இயங்குகிறது.

இவை பொதுவான சொற்றொடர்களையும், கட்டளை, வினா, வியப்பு வகைச் சொற்றொடர்களையும் மொழி பெயர்க்க வல்லது.

30க்கும் மேற்பட்ட பெயர், வினைச் சொல்லுருபுகளையும் இது கொண்டிருக்கிறது.

எடுத்துக்காட்டு

Tamil Virtual University is in Chennai என்று தட்டச்சுச் செய்தால், தமிழ் இணையப் பல்கலைக் கழகம் சென்னையில் உள்ளது என்ற மொழிபெயர்ப்பு உடனே கிடைத்து விடும்.

ஆனால் நீண்ட சொற்றொடர்களை மொழிபெயர்க்க, மேலும் தமிழ்ச் சொற்களையும், இலக்கணக் கூறுகளையும் இம்மென்பொருளில் சேர்த்து மேம்படுத்த வேண்டும்.

இம் மேம்பாட்டுக்கு வழிகோலும் முயற்சியாக தமிழ் 'சொல்வலை' ஒன்றை உருவாக்கும் திட்டமும் செயல்படுத்தப்பட்டது.

4.4 தமிழ்சொல்வலை உருவாக்குதல் (Tamil Word Net)

இது ஒரு தமிழ்ச் சொல்லில் உள்ள உருபங்களைப் பிரித்து ஆராயும் திறன் கொண்டது. நூல்கள் என்ற ஒரு சொல்லை நூல் + கள் என்று இரண்டு உருபங்களாகப் பிரிக்கும்.

மேலும் ஒரு சொல்லுக்கு ஈடான பல சொற்களையும், பல பொருள்களையும் தரவல்லது. எடுத்துக் காட்டாக, படி என்ற சொல்லுக்கு நான்கு பெயர்ச்சொல் பொருள்கள் உள்ளன. அவை,

படி - படிக்கட்டு, நாட்டளகு, வழி, வரவு செலவு

படி என்ற சொல்லுக்கு 2 வினைச் சொல் பொருள்கள் உள்ளன. அவை,

படி - படித்தல், படிதல்

அகராதி முறையில், இணையத்தில் சொற்களைப் பார்க்கும் முறையில் இது அமைக்கப் பட்டிருக்கிறது.

4.5 மின் அகராதித் தகவல் தளம்

அகராதி தொடர்பான ஒரு தகவல் தளத்தை உருவாக்குவதே இம்மென்பொருளின் நோக்கமாகும். ஒரு சொல்லுக்கான பொருளைப் பெறும் வழக்கமான தேவைகளைத் தவிர, உருபன் ஆய்வு, சொல் அடிப்படையிலான அடுக்குகளைப் பெறுதல், ஒரு சொல்லுக்கான ஒலிப்பைப் பெறுதல், ஒரு பெயர் அல்லது வினைச் சொற்களைப் பெறுதல் போன்ற பல்வேறு பயன்களும் இதில் உண்டு.

எடுத்துக்காட்டு: சொல், திரும்பு.

இலக்கண விவரம்	பொருள் (Meaning)	ஒருபொருள் பன்மொழி (Synonym)	சொல்லுக்கு (paradigms)
திரும்பு - வினை (intramitive)	return, turn to one side பழைய நிலைக்கு வருதல்	திரும்பி	திரும்புங்கள் திரும்புக
திரும்பு - வினை (finite)	அல்லது செல்லுதல் வேறு திசையை நோக்கி நகர்தல்.		திரும்பட்டும் திரும்பலாம்.

4.6 சொற்களஞ்சியம் ஆய்வுக் கருவிகள் (Corpus Analysis Tools)

தமிழில் பெயர், வினை, பெயரடை, வினையடை ஆகிய நான்கு வகைச் சொற்களையும் பிரித்து உருபங்களை இம்மென்பொருள் தனித்தனியே எடுத்துக் கொடுக்கும். சுமார் 2000 சொற்களுக்கு இது செயல்படுத்திப் பார்க்கப்பட்டது. இதற்காகத் தயாரிக்கப்பட்ட தரவுகளில் முக்கியமானது தமிழ் இலக்கண விதிகளாகும். உருபன் பிரிப்பான், உருபங்களாகப் பிரிப்பதோடு, அவற்றின் இலக்கண வகைப் பாட்டையும் காட்டும். மேலும் இதனால் ர - ற, ல - ள, ந-ண-ன எழுத்துக்களால் உண்டாகும் சொல் மயக்கங்களையும், வேறு காரணங்களால் ஏற்படும் பொருள் மயக்கத்தைத் தீர்க்கவும் முயற்சிகள் மேற்கொள்ளப்பட்டுள்ளன.

4.7 பேச்சிலிருந்து எழுத்துக்கான மென்பொருள்

பேச்சைப் புரிந்து கொண்டு தட்டச்சு வடிவில் தருவதே இதன் நோக்கமாகும். இது 5000 சொற்களையும், சொற்றொடர்களையும் மாற்ற வல்லது. இம்மென்பொருள் வேளாண்மை என்னும் ஒரு குறிப்பிட்ட துறையில் செயல்படுத்தப்பட்டது.

எடுத்துக்காட்டாக, ஒரு விவசாயி கூறுவதை இம்மென்பொருளால் தட்டச்சுச் செய்ய முடியும்.

4.8 கையெழுத்தைத் தட்டச்சாக மாற்றும் மென்பொருள்

இம்மென்பொருள் கையெழுத்துப் படிவங்களைப் படித்து, எழுத்துக்களை அறிந்து கொண்டு அவற்றைத் தட்டச்சு வடிவில் தரவல்லது. இது யாருடைய கையெழுத்தையும் புரிந்து கொள்ள வல்லது.

ஒரு பக்கத்தில் உள்ள கையெழுத்துப் படிவத்தின் ஒரு பகுதியையோ, படத்தையோ கூட இதனால் தட்டச்சு வடிவில் தர முடியும்.

4.9 செல்பேசிகளில் தமிழில் தகவல் அளித்தல்

அனைத்துச் செல்பேசிகளிலும், தொடர்வண்டி, வானூர்தி, தொடர்பான நேர அட்டவணை, பதிவு விவரம் ஆகிய விவரங்களை இன்று ஆங்கிலத்தில் பெறுவதைப் போல, தமிழிலும் பெறலாம். மேலும் பல்வேறு நிறுவனங்களின் தற்போதைய பங்கு பற்றிய தகவல்களையும் பெறலாம்.

4.10 செல்பேசிகளில் தமிழ்

செல்பேசிகளில் தமிழில், குறுந்தகவல்களை அளித்தல், பெயர்களைப் பதிவு செய்தல் போன்றவற்றையும் செய்யும் மென்பொருள் இது. இதற்கான விசைப்பலகை வடிவமைப்பும் இதில் செய்யப்பட்டுள்ளது. இம்மென்பொருளின் அடுத்த கட்டமாக, முதல் எழுத்தை உள்ளீடு செய்யும் போது அந்த எழுத்தில் தொடங்கும் சொற்களையும் கணித்துத் தரவல்ல மென்பொருளும் இப்போது உருவாக்கி கொண்டிருக்கிறது.

4.11 கணிப்பொறிகளில் தமிழ்

மைக்ரோசாப்ட், லினக்ஸ், போன்ற செயலிகளை ஆங்கிலம் தெரியாதவர்களும், தமிழ், கொண்டே பயன்படுத்தும் பல மென்பொருள்கள் உருவாக்கப்பட்டுள்ளன.

5.0 மென்பொருள் உருவாக்கச் சிக்கல்கள்

ஏறக்குறைய எல்லா மென்பொருள் திட்டங்களும் கொடுத்த காலத் தவணைக்குள் முடிக்காமல் பல ஆண்டுகள் கழித்தே முடிக்கப்பட்டிருக்கின்றன. பல 10 ஆண்டுகள் ஆகியும் முடிக்கப்படாமல் பல்வேறு நிலைகளில் தேக்கமடைந்து உள்ளன. இதற்கு மென்பொருள்களை உருவாக்கும் பொறியாளர்கள் இடப்பெயர்வும், குறைபாடுகளைப் புரிந்து கொண்டு களையும் திறமையில்லா புதியவர்களும், உருவாக்கப்படும் மென்பொருள்களுக்குச் சந்தையில் வரவேற்பு இன்மையும் காரணங்களாகும்.

முடிக்கப்பட்டு, செயல்முறையில் ஒரு கணிப்பொறியில் காட்டப்படும் சில மென்பொருள்களும், வேறு சில கணிப்பொறிகளில் செயல்படுவதில்லை. இம்மென்பொருள்கள் கணிப்பொறி, காலம், மொழி என்ற வேறுபாடுகளால் செயல்படாமல் போய்விடுகின்றன. எனவே, உருவாக்கப்படும் பல மென்பொருள்கள் வெளிச்சத்துக்கு வராமல் முடங்கிக் கிடக்கின்றன.

6.0 முடிவுரை

தமிழ் மென்பொருள்களின் தேவையை உணர்ந்து பரவலாக்கினால், அவை முழுமையடைந்து சந்தைக்கு வர வாய்ப்பிருக்கிறது. அதற்கு அரசு நிர்வாகத்திலும், கல்வியிலும், நீதிமன்றங்களிலும் தமிழ் முதன்மை பெற வேண்டும்; தமிழ்நாட்டில் தமிழ் ஆட்சி செய்ய வேண்டும். அந்நாளை எதிர்நோக்கிக் காத்திருப்போம்.

Moodle: For Enhanced learning (Tamil language)

Ravishankar Somasundaram
Raviepic3@gmail.com

Introduction

As we all know “The Webolution(web-evolution)” has changed the way we live. How many of us still have the fear of getting lost without knowing our way in some foreign region?

From house hold utilities to online traffic monitoring systems the web has its part to play, to mention about the way it has impacted the learning aspects, companies foresee a save on range of millions bringing in the e-learning concept into their environment.

Having worked on the lines of mashing up cutting edge technologies along with learning aspects to provide e-learning solutions, I suggest Moodle – one of the best available in market today to deliver e-learning content.

What do I gain on reading further?

Through this paper you will get to know **what** an LMS is and **why** do we need to adopt one **for learning Tamil language**. We will also see **how** can it **make a difference** in today’s conventional way of learning and a combination of **why moodle** is the best choice, **what can it offer**, **what moodle cannot offer** and **how can we compensate it** to bring in **maximum learning effectiveness** by employing featured customizations based on research on this platform in a way it suits best for teacher and student.

Aiming the applications of this platform in **Tamil language teaching**, Have created a demo course in Tamil language which illustrates in detail the usage of all features (types of resources and activities) which can be utilized within a course to form course content.

The course can be accessed under Tamil category in this site (<http://demo.moodle.net/login/index.php>); having covered the existing features of moodle by other presenters as well the demo course mentioned above, this paper is primarily going to **focus on the featured customizations to enhance the learning**

- **Curriculum Management**

Necessity: To avoid the student turning aimless and cruise within the site as no boundaries/learning objectives has been set for him to progress in achieving his goal.

Features: Creating and maintaining course, module hierarchy, Course completion methodologies and roles associated with the curriculum.

- **Learning Effectiveness**

Necessity: To help, motivate, monitor the student progress along with their peers.

Features: Includes Learning Curve mapping and Statistics which results in a healthy competition with their fellow students on comparing their grades.

What is an LMS?

LMS = Learning Management system, as its name indicates a software application for managing, administering, documenting and tracking learning content which involves students and teachers with respect to learning objectives.

In other words it's a software package which delivers a new and effective methodology of teaching and learning for both students and teachers via network or as a standalone pack

Why do we need an LMS for learning Tamil language?

Because a system which can make a global reach is need of this hour, using an LMS we

- Open up a possibility of working with people who aren't physically with you either sometimes or all of the time.
- Help students who are studying in a typical class room environment this by providing a continuous learning opportunity using which he can not only be in touch with the learning concepts studied in the class but actually enhance it by discussing, clarifying his doubts, answering others doubts and by taking part in other forms of activities with or without the help of resources provided within the LMS.
- Uniting people all over the world in the name of Tamil by providing a channel for communication between people, who are
 - Experts in Tamil language
 - Dedicated towards learning Tamil language
 - Interested to know about Tamil language

Irrespective of their geographical location.

How can it bring a difference?

It can have a tremendous impact because of the learning methodology and flexibility it provides to students and facilitators. Have quoted few

Students

- Provides the flexibility for students to attend the course in their desired timings.
- Students have a reach to the teacher/facilitator even if they are not online.
- Majorly it not only makes students familiar with computer and web portal usage but also inculcates the blended learning approach and community discussion qualities which are a must to have qualities on higher stages of their career.

Teacher/Facilitator:

- Accommodate more students than it is possible in a typical class room setup
- Opens up possibility on conducting events like "Experts corner, Facilitators week . . . etc" which needs dedicated time of a guest sharing real time situations with students which becomes often less possible in a class room setup because of the necessity of guests physical presence amongst their varying busy schedule.

- Single repository to store all course related materials thus preventing deterioration with time, and making them available through ages.

Overall the throughput of students as well as teacher is high due to the flexibility an LMS adds up to their learning environment.

Why moodle?

Strong user base: Moodle is established in 206 countries, its count on number of registered sites is 46,995 which gives the **reliability** that the future development of moodle is no way diminishing.

Healthy support: A software package no matter to what extent it can outperform others in functionality and other aspects, without a manual it is used less/useless. Moodle not only have a vast group of individuals who regularly contribute to **documentation** and try to keep it **up to date**, what makes moodle special is the **densely populated forums** where you get answers for any question from co moodlers almost immediately.

Open source: As familiarly known moodles source code is freely available **under GPL license**, in case of doubts regarding moodle becoming proprietary or getting sold to some X company in future, here(<http://docs.moodle.org/en/Future>) is a comprehensive material for you.

What can moodle offer?

Moodle can offer a **perfect teaching and learning environment** for both students and teachers along with **powerful mechanisms to validate the theoretical as well as implied, applied learning** acquired by the students.

What moodle does not offer as of now?

Present version of moodle does not offer the following important features:

Mapping the learning curve attained by the student and formulate a focused attention to required students on required areas.

As a facilitator it is our responsibility to indicate it to the students their strengths and weakness.

For example, Theoretical knowledge and applied knowledge.

We need to measure and show the student whether he is good in applying the concepts learned in the classroom. In many cases we notice that a student has good knowledge on theory but fails to apply on real time situations, hence focus has to be given on that phase for that particular student.

Providing a statistics on what is happening around him in the online class room with respect to learning.

Because the students are getting trained in a VLE (Virtual Learning Environment), they lose the opportunity to have a healthy competition with their fellow students by comparing their grades.

Thereby not knowing their stance in the class, this can diminish the interest within students on learning online on due course of time.

Manage the flow between constituents of a course and between courses themselves to form a curriculum.

To emphasize the necessity and importance of a curriculum management system consider this scenario

A portal which is dedicated for learning Tamil language, containing different courses starting from basics of Tamil (alphabets) to writing a comprehension. Now a student enters the site with an aim to attain expertise in Tamil, and all he sees is multiple courses, now here the student

1. Can skip one or multiple mandatory courses knowingly or unknowingly.
2. Can jump resources/activities/courses as a learning structure doesn't restrict him in doing so.
3. Gets lost often when he logs into the site without knowing where he left last time.
4. Might not know when he has to go to next activity or resource/activity/course.

Solution?

Developing a solution with keeping the above issues in mind along with usability factor was challenging.

Curriculum Management:

Necessity: To avoid the student turning aimless and cruise within the site as no boundaries/learning objectives has been set for him to progress in achieving his goal.

Add-on: Curriculum functionality was designed, implemented, and was subjected to all kinds of tests; the final product has been contributed in moodles contrib section. You can find it in here (<http://tracker.moodle.org/browse/CONTRIB-1604>).

Presently this functionality is up and running on a site which was built to handle 3000 concurrent users and 20,000 users overall using (Cloud computing - Amazon EC2).

Features:

Creating and maintaining courses along with module hierarchy, Course completion methodologies and roles associated with the curriculum.

Course hierarchy

1. There can be a tree hierarchy
2. There can be a parallel hierarchy
3. There can be a serial hierarchy

Between courses inside the curriculum, the admin has the power to tailor the courses in any of above mentioned hierarchies.

Module hierarchy

1. There can be a tree hierarchy
2. There can be a parallel hierarchy
3. There can be a serial hierarchy

Inside a course or a curriculum, the admin/teacher has the power to tailor the activity/resource flow in any of above mentioned hierarchies. If a course/curriculum has elements following any of the above mentioned hierarchy, it is said to be possessing dependencies.

Course completion: Completion of a course for any student occurs on two ways,

Automatic: Any student is marked as he completed the course if he completes the dependencies within the course; once he completes this course automatically he gets access to other courses which are dependent on this course.

Manual: A student takes a course and even though he finished all locks/dependencies he will not be marked as he completed the course until the teacher manually specifies that he/she is qualified enough to move to further. And any student enrolled in a curriculum is enrolled in all the courses within the curriculum but denied access to courses as per the hierarchy structure designed by admin/teacher. Teacher/admin has the facility to suspend the user for any single course or multiple courses in case he violates some rules for that course/curriculum itself.

Roles:

- 1. Curriculum wide (similar to site wide roles)
- 2. Course wide (similar to course wide)

Learning Effectiveness:

Learning Curve mapping:

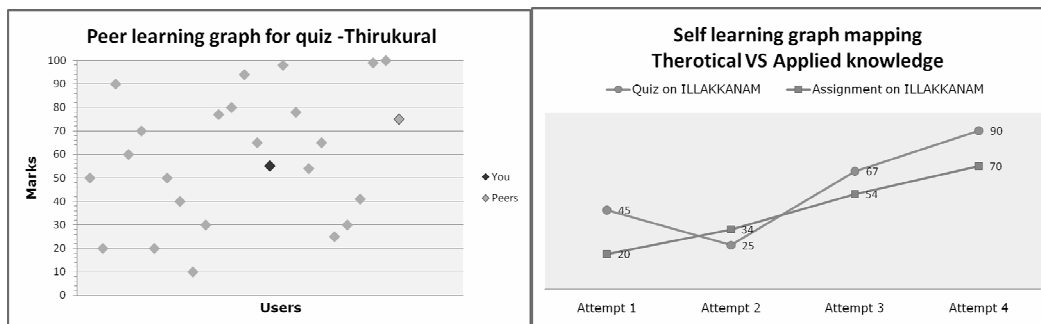
Necessity: To help, motivate, monitor the student progress along with their peers.

Add-on: This functionality was designed, implemented, and was subjected to all kinds of test's, the final product has been implemented in sites where users from diverse backgrounds and job natures find it impeccably meeting the motto which is to help, motivate, monitor the student progress along with their peers and create a healthy competition between them.

Features: Includes Learning Curve mapping and Statistics which results in a healthy competition with their fellow students on comparing their grades

Statistics:

Indicating the current position of the student amongst his peers using discrete scatter plot.



Indicating the learning curve of required activities for a student, quiz mapping to get the learning curve on theoretical knowledge and Assignment mapping to get the learning curve on applied knowledge using continuous line plot.

Quality Analysis of Tamil Virtual University

S.Rajkumar

ME Industrial engineering

Kumaraguru engineering college - Coimbatore

Srajkumar24@yahoo.co.in

Abstract

This document summarizes the findings of an independent study concerning online education, virtual universities and Tamil language instruction being conducted over the Internet. Widespread concern about the erosion of Tamil identity and decay of Tamil language has prompted officials of the State Government of Tamil Nadu to take prompt steps to address the issues. In early 1999 the State Government announced its intention to establish a **Tamil Virtual University** designed to promote Tamil language, literature and culture internationally through the medium of Internet-linked computers. A High-Level Committee consisting of senior Tamil educators has been formed and sub-committees drawn up with a mandate to formulate the vision and mission of the Tamil Virtual University, issue a detailed report to the State Government with recommendations. Thus we are going to analyse what exactly people needs and requirements through a technique called Quality Function Deployment (QFD). QFD is a methodology for incorporating the Voice of the People(VOP) into the practical design. It aims to capture and prioritise people requirements and translate them into design requirements through the use of management and planning tools such as affinity diagrams, tree (hierarchy) diagrams, relations diagrams, matrices and tables.

QFD can be used to analyse the TVU such as:

- How can we extract people requirements and prioritise them?
- How can we identify the key design requirements that will help satisfy our people?
- Which people requirements should we be focusing?

1. Understanding people Requirements
2. Quality Systems Thinking + Psychology + Knowledge/Epistemology
3. Maximizing Positive Quality That Adds Value
4. Comprehensive Quality System for Satisfaction

Thus the Tamil Virtual University encloses:

- ❖ Understanding '**true**' **needs** from the people perspective
- ❖ What 'value' means to the people, from the people perspective
- ❖ Understanding how people or end users become interested, choose, and are satisfied

- ❖ Analyzing how do we know the needs of the people
- ❖ Deciding what features to include
- ❖ Determining what **level of performance** to deliver
- ❖ Intelligently linking the needs of the people with design, development, engineering, manufacturing, and service functions .
- ❖ Intelligently linking Design with the front end Voice of people analysis and the entire design

The Tamil Virtual University **website should include:**

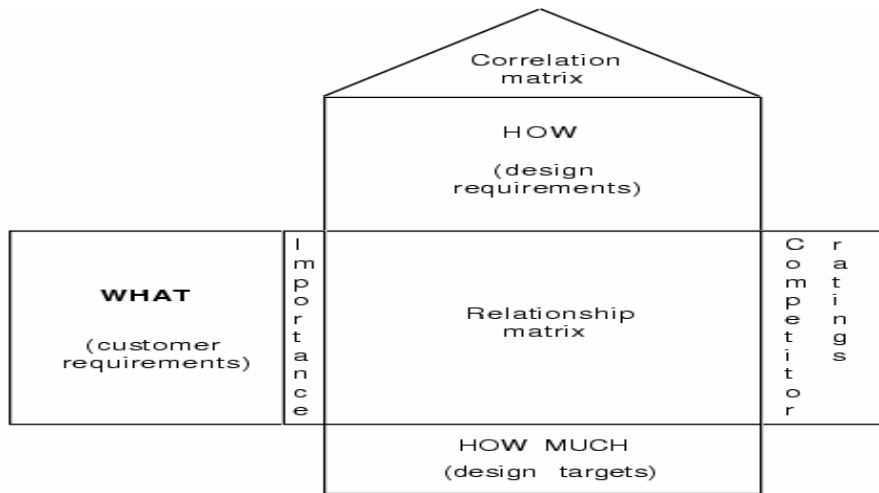
- ❖ a virtual campus 'map' of programs and offerings on its home page;
- ❖ an attractive TVU logo or banner on each TVU Web page;
- ❖ a complete listing of course offerings at Tamil Nadu universities;
- ❖ information about **admission to institutions** of higher learning in Tamil Nadu; and
- ❖ A listing of faculty and staff engaged in the TVU project.

Objectives of the analysis:

- ❖ To deliver online learning to geographically separate Tamil communities;
- ❖ To deliver customized programmes for the promotion of Tamil language and culture;
- ❖ To develop educational courses to address the needs of Tamils living abroad;

Mission

1. To develop and deliver Internet based learning material in Tamil language, literature and culture to global Tamil Communities and others interested.
2. To initiate and continue necessary measures to co-ordinate and pool together knowledge resources, developed in Tamil in different parts of the world, for wider dissemination.



TVU ANALYSIS TABLE

	Material	Clear to Point	Comfort View	Quality	Satisfaction	Service Ability	Screen Complex	Design Aspects	Look and feel	Appearance	Volume	Convenience	Availability	Return visit rate	Targeting	Performance	Target value		
Audio & video resources	▲	△	▲	●	●	▲	▲	△	△	▲	△	▲	▲	▲	▲	3	4		
e-texts	▲	▲		△	▲	●		▲		△		▲	▲	▲		1	4		
People feedback			△	●		●			△					▲		2	5		
helpline				△	▲	●	▲		△			▲	▲		△	2	5		
Message board		●	●	△	▲		●	▲		△	△	▲		△	△	1	5		
Comment box			●	●	△				△			▲		▲		2	4		
quality	●	●	●	●	▲	●			△			▲	●		△	●	3	5	
hyperlinks	▲	●		▲	▲				▲		△		▲				2	4	
Glossary			▲	▲		▲	▲	△		△	△		▲		△		2	4	
Course pack	▲	▲		▲		▲	▲			●		▲		▲		▲	3	5	
Library resources	▲			▲	●		△	△	▲	●	△		▲	▲	▲		3	4	
Books collections	▲	▲		△	△	●	▲	△	▲		▲	△	▲	▲			3	4	
Essay collection		▲	▲	△		●		●		●	▲		△	▲	▲	●	1	4	
Poetry collection	▲	▲		▲	△	▲	▲	△		▲		▲		△		▲	1	4	
History details	▲				●	▲	△			●	▲		▲	▲	△		●	1	4
Down loading options					●	▲	△				▲	▲	▲		▲		▲	1	3
Remain interested	▲		▲	▲		●		▲	▲		△		▲	△	△	△	△	2	5
Casual environment		▲	▲	▲		△	△		▲				●	△	●		●	2	4
Dates and calendar		△		▲	△		●			●	△		△		▲			1	3
Tvu logo and banners	▲			△	△	△		▲	△	▲	●				△	●		2	4
Tvu project details	▲	▲		▲		●	●	△		▲	△		▲		▲		●	3	4
Campus map		▲			△		●		▲	▲					△	▲		3	4
Press releases		▲		▲	△	△	●		△		●		△	●		▲		1	3
Recent visited blogs			▲	△		●	△		●		●	●		△		△	●	1	3
performance	4	3	3	3	3	3	2	3	2	3	2	3	2	2	2	3	2	3	
target	5	4	4	4	4	5	5	4	4	4	4	4	4	5	5				

Steps in QFD in TVU:

- ❖ Plan collection of people needs.
- ❖ Prepare for collection of people needs. Identify required information. Prepare agendas, list of questions, survey forms, focus group/user meeting presentations
- ❖ Determine people needs or requirements. Document these needs. Consider recording any meetings. Extract statements of needs from documents. Summarize surveys and other data. Use techniques such as ranking, rating, paired comparisons, or conjoint analysis to determine importance of people needs.
- ❖ Use affinity diagrams to organize people needs. Consolidate similar needs and restate. Organize needs into categories. Breakdown general people needs into more specific needs by probing what is needed. Once needs are summarized, consider whether to get further people feedback on priorities. Undertake meetings, surveys, focus groups, etc. to get people priorities. State people priorities using a 1 to 5 rating. Use ranking techniques and paired comparisons to develop priorities.
 - - strong(5) ♠ - medium (3) △ - weak(2)

TVU Planning:

- ❖ Organize people needs in the Product Planning Matrix. Group under logical categories as determined with affinity diagramming.
- ❖ Establish critical internal people needs or management control requirements;
- ❖ State people priorities. Use a 1 to 5 rating. Critical internal people needs or Develop competitive evaluation of current TVU. Use surveys, people meetings or focus groups/clinics to obtain feedback. Rate scale with "5" indicating that the TVU fully satisfies the people needs.
- ❖ Review the competitive evaluation strengths and weaknesses relative to the people priorities. Determine the improvement goals and the general strategy for responding to each people need. The Improvement Factor is "1" if there are no planned improvements to the competitive evaluation level. Add a factor of .1 for every planned step of improvement in the competitive rating, (e.g., a planned improvement of going from a rating of "2" to "4" would result in an improvement factor of "1.2").
- ❖ The process of setting improvement goals and sales points implicitly develops a product strategy. Formally describe that strategy in a narrative form. This strategy brief is typically one page and is used to gain initial focus within the team as well as communicate and gain concurrence from people improvement factor, and the weighting factor associated with the relationship in each box of the matrix

Conclusion:

Thus QFD is a methodology for incorporating the Voice of the People (VOP) helped to develop content that are attractive and suit the needs of young people living in foreign cultures, they will frequent the TVU Website and make use of its course offerings as auditors at first and deliver customised programmes to meet the cultural needs of the Tamil Communities in different parts of the world and help them retain contact with their heritage. So that to facilitate easy access to other online resources already developed by the international Tamil communities.

மலேசியாவில் தமிழ்க்கல்வியின் தொடர்ச்சியும் வளர்ச்சியும்

மழலையர்ப் பள்ளி முதல் பல்கலைக்கழகம் வரை

மன்னர் மன்னன் மருதை
மொழி-மொழியியல்புலம்,
மலாயாப் பல்கலைக்கழகம்

தோற்றுவாய்

தமிழகத்திற்கு வெளியே பல நாடுகளில் தமிழ்மொழிப் பயன்பாடு மெச்சத்தக்க வகையில் உள்ளது. ஈழத்தை அடுத்து, மலேசியாவில் தமிழ்க்கல்வி மிகவும் உன்னத நிலையில் உள்ளது. இக்கட்டுரை, மலேசியாவில் மழலையர்ப் பள்ளி முதல் உயர்க்கல்விக் கூடம் வரை தமிழ்க்கல்வியின் வளர்ச்சி மற்றும் தொடர்ச்சி பற்றி விவரிக்கின்றது. மலேசியாவில் வாழ்கின்ற 20இலட்சம் இந்தியர்களில் 85 விழுக்காட்டினர் தமிழர்களாவர். 1870-களில் பிரிட்டிஷ் காலனித்துவ ஆட்சியாளர்களால் தமிழர்கள் இந்நாட்டிற்குக் கொண்டுவரப்பட்டனர். சஞ்சிக்கூலிகளாக இந்நாட்டிற்குக் கொண்டுவரப்பட்ட தமிழர்கள் தங்களுடைய மொழியையும் பண்பாட்டையும் பேணிக்காப்பதில் பல்வேறு போராட்டங்களை நடத்தி வந்துள்ளனர்.

காலனித்துவ ஆட்சியாளர்களால் 1912-இல் நிறைவேற்றப்பட்ட தொழிலாளர் சட்டம், தோட்டத் தொழிலாளர்களின் பிள்ளைகளுக்குக் கல்வி வழங்க தோட்ட நிர்வாகங்களை வற்புறுத்தியது. இதன்வழி நாடு முழுவதும் உள்ள தோட்டங்களில் தமிழ்ப்பள்ளிகள் அமைக்கப்பட்டன. இதன்வழி, தமிழ்ப்பள்ளிகளின் எண்ணிக்கையும் துரித வளர்ச்சிக் கண்டது. 1930-இல் 333-ஆக இருந்த எண்ணிக்கை 1938-இல் 547 ஆகவும், 1947-இல் 741 ஆகவும், 1957-இல் 888 ஆகவும் உயர்வு கண்டது.

1959களில் ஏழாம் வகுப்புவரை தமிழ் படிப்பதற்கான வாய்ப்பு இருந்தது. தொடர்ந்து அமலாக்கப்பட்ட பல்வேறு கல்விச்சட்டங்கள் LCE (9ஆம் வகுப்பு), SC/MCE (11ஆம் வகுப்பு), HSC (12ஆம் வகுப்பு) ஆகிய நிலைகளில் தமிழ்மொழியைத் தேர்வுப் பாடமாக எடுத்துப் பயில வழிவகுத்தன. இவ்வளர்ச்சியானது தொடர்ந்து பல்கலைக்கழகத்திலும் தமிழ் பயில வாய்ப்பு ஏற்படுவதற்கான உந்துதலை ஏற்படுத்தியது.

மழலையர்ப் பள்ளி

மழலையர்ப் பள்ளிகள் பல ஆண்டுகளாக இந்நாட்டில் இயங்கி வருகின்றன. மழலையர்ப் பள்ளிகளில் நீண்ட காலமாக ஆங்கிலமும் மலாய்மொழியும் பயிற்றுமொழியாக இருந்து வந்தன. 1985 முதல் தமிழைப் பயிற்றுமொழியாகக் கொண்ட சில தனியார் மழலையர்ப் பள்ளிகள் அமைக்கப்பட்டன. தொடர்ந்து முறையான பாடத்திட்டத்தைக் கொண்டு அரசாங்கத்தால் ஒரு சில தமிழ்ப்பள்ளிகளில் மழலையர்ப் பள்ளிகள் அமைக்கப்பட்டன.

மழலையர்ப் பள்ளியின் அவசியத்தையும் இன்றியமையாமையும் கருத்தில் கொண்டு இன்று நாடு முழுமையும் 163 தனியார் தமிழ் மழலையர் வகுப்புகள் இயங்கி வருகின்றன. இத்தனியார் தமிழ் மழலையர் வகுப்புகளைத் தமிழ் அறவாரியம், இந்து சங்கம், ஆலய நிர்வாகங்கள், பெற்றோர் ஆசிரியர் சங்கம் போன்ற ஒன்பது பொது அமைப்புகள் இணைந்து நடத்தி வருகின்றன என்பது குறிப்பிடத்தக்கது. இவ்வகுப்புகளில் 4950 குழந்தைகள் கல்வி கற்று வருகின்றனர்.

கல்வி அமைச்சால், நாடு முழுமையும் உள்ள சில தமிழ்ப் பள்ளிகளில் 158 தமிழ் மழலையர் வகுப்புகள் நடத்தப்பட்டு வருகின்றன. இவ்வகுப்புகளில் 3950 குழந்தைகள் கல்வி பயின்று வருகின்றனர். கல்வி

அமைச்சின், ஆசிரியர் கல்விக் கழகங்களில் முறையான பயிற்சி பெற்ற தமிழ் ஆசிரியர்கள் இவ்வகுப்புகளில் கற்பித்தலைத் திறம்பட மேற்கொண்டு வருகின்றனர்.

தமிழ் தொடக்கப் பள்ளிகள்

மலேசிய நாட்டில் முறைசார்ந்த தொடக்க நிலைத் தமிழ்க்கல்வியை வழங்குவதில் தமிழ் தொடக்கப் பள்ளிகள் முன்னிலை வகிக்கின்றன. ரெவரெண்ட் ஹட்சிங்ஸ் (Rev. R. Hutchings) அவர்களால் 1816இல் பினாங்கில் ஃபிரி ஸ்கூலின் (Penang Free School) ஒரு பிரிவாகத் தமிழ் வகுப்பு தொடங்கப்பட்டது. இதுவே, இந்நாட்டின் முதல் தமிழ்ப்பள்ளியாகும்.

இந்நாட்டின் தேயிலை, காப்பி, கரும்பு, தென்னை, ரப்பர் தோட்டங்களில் வேலை செய்ய 19ஆம் நூற்றாண்டின் பிற்பகுதியில் ஏராளமான தொழிலாளர்கள் தென்னிந்தியாவிலிருந்து ஆங்கிலேயர்களால் இங்கு வரவழைக்கப்பட்டனர். இத்தொழிலாளர்களை இங்கேயே தொடர்ந்து தங்கவைப்பதற்குப் பிரிட்டிஷார் வியூகம் வகுத்தனர். அவ்வகையில் தொழிலாளத் தமிழர்கள் தங்குவதற்குத் தோட்டத்திலேயே வீடு, ஆயாக் கொட்டகை, கள்ளுக் கடை, முனியாண்டி/மதுரைவீரன் கோயில், தமிழ்ப் பள்ளிக்கூடம் என அமைத்துத் தந்தனர்.

பிரிட்டிஷாரின் பிரித்தானும் கொள்கை மற்றும் சுயநலத்தன்மையால் தொடங்கப் பெற்ற தமிழ்ப்பள்ளிகள் நெடுக பல்வேறு சிக்கல்களையும் மாற்றத்தையும் கண்டு வந்துள்ளன. அதோடு, அரசாங்கத்தின் கல்விச்சட்ட மாற்றங்களுக்கு உட்பட்டு நல்ல வளர்ச்சி நிலையினை அடைந்துள்ளது குறிப்பிடத்தக்கது. இன்றைய நிலையில் மலேசியத் தமிழர் வாழ்வில் தமிழ்ப்பள்ளிகள் இரண்டறக் கலந்துள்ளது என்பது வெள்ளிடைமலை.

இப்பொழுதைய நிலவரப்படி நாடு முழுமையும் 523 தமிழ் தொடக்கப் பள்ளிகள் சிறப்பாக இயங்கி வருகின்றன. பள்ளிகளின் எண்ணிக்கை குறையினும், மாணவர்களின் எண்ணிக்கை ஆண்டுக்கு ஆண்டு உயர்வு கண்டு வருகின்றது. 2000ஆம் ஆண்டில் 89,175 ஆக இருந்த தமிழ் தொடக்கப் பள்ளிகளின். மாணவர் எண்ணிக்கை 2005ஆம் ஆண்டில் 98,579ஆக உயர்வு கண்டது. 2010ஆம் ஆண்டில் அந்த எண்ணிக்கை 105,000ஆக உயர்வு கண்டுள்ளது குறிப்பிடத்தக்கது. 2010இல் முதலாம் வகுப்பில் சேர்ந்த மாணவர் எண்ணிக்கை 17,650ஆகும். அதேவேளையில் தமிழ் தொடக்கப் பள்ளிகளில் பணிபுரியும் ஆசிரியர்களின் எண்ணிக்கை 8700 ஆகும்.

மலேசியாவில் உள்ள மிகப்பெரிய தமிழ்ப்பள்ளி, சிலாங்கூர் மாநிலத்தின் கிள்ளான் பட்டணத்திலுள்ள சிம்பாங் லீமா தமிழ்ப்பள்ளியாகும். இப்பள்ளியில் 2264 மாணவர்கள் தமிழ்க்கல்வி பயின்று வருகின்றனர். இப்பள்ளியில் 106 ஆசிரியர்கள் பணிபுரிகின்றனர். இவ்வாண்டு முதலாம் ஆண்டில் மட்டும் 11 வகுப்புகளில் 400 மாணவர்கள் சேர்ந்துள்ளனர். நல்ல நிலையில் உள்ள விரிவுரையாளர், வழக்கறிஞர், மருத்துவர், ஆசிரியர், வணிகர் போன்ற பலதுறைகளைச் சார்ந்த பெற்றோர்கள் தங்களது பிள்ளைகளைத் தமிழ்ப் பள்ளிகளுக்கு அனுப்பி வருகின்றனர்.

தேசியப் பள்ளிகளில் தமிழ்க்கல்வி

தேசியப்பள்ளிகளில் பயிலும் இந்திய, சீன மாணவர்களுக்குக் குறைந்தது 15 பெற்றோரின் விருப்பத்தின் பேரில் தாய்மொழிக் கல்வி கற்பிக்க, கல்வி பெற 1961ஆம் ஆண்டு கல்விச்சட்டம் வழிவகுத்தது. இது மாணவர்களின் தாய்மொழிக் கல்வி (Pupils' Own Language) என அழைக்கப்படுகின்றது. இச்சட்டத்தின் வழி தொடக்கப் பள்ளிகளில் மூன்றாம் ஆண்டிலிருந்து ஆறாம் ஆண்டு வரையில் ஒரு கிழமைக்கு 120மணித்துளிகள் தாய்மொழிக் கல்வி பயிலலாம். தமிழ் ஆசிரியர் பயிற்சி பெற்ற ஆசிரியர்களைக் கொண்ட பள்ளிகளில் பாட நேரத்திலும், அல்லாத பள்ளிகளில் பகுதி நேரமாகவும் தாய்மொழிக் கல்வி வகுப்புகள் தற்பொழுது நடைபெற்று வருகின்றன.

2000ஆம் ஆண்டு முதல் இந்திய மாணவர்களுக்கு மட்டுமல்லாது, தமிழைத் தாய்மொழியாகக் கொண்டிராத எல்லா இன மாணவர்களும் தமிழைக் கற்கும் வாய்ப்பினை கல்வி அமைச்சு ஏற்படுத்தித் தந்துள்ளது. ஒருசில பள்ளிகளில், கிழமைக்கு இரண்டு பாட வேளை தமிழ் கற்பிக்கப்பட்டது. ஆனால் இத்திட்டம் எதிர்பார்த்த வெற்றியை அடையவில்லை. 2007ஆம் ஆண்டு சோதனை அடிப்படையில் 70 தேசியப்பள்ளிகளில் தமிழ் கற்பித்தல் அறிமுகப்படுத்தப்பட்டது. இவ்வாண்டு நாடு முழுமையும் 200 தேசியப்பள்ளிகளில் இத்திட்டம் நடைமுறைப்படுத்தப் பட்டுள்ளது.

இடைநிலைப்பள்ளிகளில் தமிழ்க்கல்வி

இடைநிலைப்பள்ளிகளில் புகுமுக வகுப்பு முதல் ஐந்தாம் படிவம் வரையில் குறைந்தது 15 பெற்றோரின் விருப்பத்தின் பேரில் தாய்மொழிக் கல்வி (Pupils' Own Language) கற்பிக்க 1961ஆம் ஆண்டு கல்விச்சட்டம் வழிவகுத்தது. புகுமுக வகுப்பிற்கு ஒரு கிழமைக்கு 160 மணித்துளிகளும், முதலாம் படிவத்திலிருந்து ஐந்தாம் படிவம் வரையிலான வகுப்புகளுக்கு 120 மணித்துளிகளும் தமிழ் மொழி கற்க ஒதுக்கப் பட்டுள்ளது. தமிழ் ஆசிரியர் பயிற்சி பெற்ற ஆசிரியர்களைக் கொண்ட பள்ளிகளில் பாட நேரத்திலும், அல்லாத பள்ளிகளில் பகுதி நேரமாகவும் தாய்மொழிக் கல்வி வகுப்புகள் நடைபெற்று வருகின்றன.

இடைநிலைப்பள்ளிகளில் மூன்றாம் படிவம் பயிலும் மாணவர்கள் கீழ் இடைநிலை மதிப்பீட்டுத் தேர்வில் (Lower Secondary Assessment) தமிழை ஒரு பாடமாக எடுக்கலாம். அதுபோன்றே ஐந்தாம் படிவம் பயிலும் மாணவர்கள் மலேசியக் கல்விச் சான்றிதழ் (Malaysian Certificate of Education) தேர்வில் தமிழ் மொழியை ஒரு பாடமாகவும், தமிழ் இலக்கியத்தை ஒரு விருப்பப் பாடமாகவும் எடுத்துத் தேர்வெழுதலாம்.

இதைத்தவிர ஆறாம் படிவம் பயிலும் மாணவர்கள் மலேசிய உயர்கல்விச் சான்றிதழ் (Higher School Certificate) தேர்வுக்காகத் தமிழைத் தேர்வுப் பாடமாக எடுக்கலாம். இத்தேர்வு இரண்டு தாள்களைக் கொண்டதாகும். தாள் 1 தமிழ்மொழி; தாள் 2 தமிழ் இலக்கியம் ஆகும். பள்ளிகளில் இப்பாடம் கற்பிக்கப்படுவதில்லை; பெரும்பாலான மாணவர்கள் தங்கள் சொந்த முயற்சியிலேயே கற்றுக் கொள்கின்றனர். இத்தேர்வில் சிறந்த புள்ளிகளைப் பெறும் மாணவர்களுக்கே பல்கலைக்கழகத்தில் தமிழ் பயில வாய்ப்பளிக்கப்படுகின்றது.

இந்திய ஆய்வியல் துறை : மலாயாப் பல்கலைக்கழகம்

மலாயாப் பல்கலைக்கழகத்தில் 1956-இல் இந்திய ஆய்வியல் துறை அமைவதற்குக் கார் செளண்டர்ஸ் ஆணையம் (1948) வழிவகுத்தது. கார் செளண்டர்ஸ் ஆணையம், தமிழ் ஆய்வுத்துறையை அமைக்க வேண்டுமெனவும், தமிழ் மட்டுமல்லாது திராவிடப் பண்பாடு, தென்னிந்திய வரலாறு ஆகியனவும் உள்ளடக்கி இருக்க வேண்டுமெனவும் பரிந்துரைத்தது.

“...We conclude therefore that a place should be found for Tamil studies. They should not be limited to Tamil language, but should include the whole range of Dravidian Culture and South Indian history.”

(Report of the commission on University Education in Malaya under the chairmanship of sir Alexander Carr- Saunders, the Government Press KL, 1948, pp43-44; quoted by Xavier S.Thani Nayagam, 1968, pp215,216, Tamil Studies Abroad a Symposium)

1946 ஆகஸ்ட் திங்கள் 3,4,5 ஆம் நாட்களில் நடைபெற்ற ம.இ.காவின் முதலாவது கூட்டத்தில் இந்தியர்களுக்கு இந்திமொழி பரிந்துரைக்கப்பட்டது. இதற்கு, சிங்கப்பூரில் பணியாற்றிய இந்தியத் தொழிலாளர்கள் கடும் எதிர்ப்பு தெரிவித்தனர். இந்தியர்களுக்கெனத் தாய்மொழி தமிழே வேண்டுமெனப் போராடினார்.

(Indian Daily Mail 19.8.1946, as quoted by Michael Stensen, p153,159; Class, Race & Colonialism in West Malaysia, 1980)

இந்திய ஆய்வியல் துறையில் சமஸ்கிருதமே முதன்மை மொழியாக இருக்கவேண்டுமென நீலகண்ட சாஸ்திரி முன்மொழிந்ததாக அறியப்படுகிறது. இதனை எதிர்த்து, தமிழ் முரசு நிறுவனர் தமிழ்வேழ கோ.சாரங்கபாணி 'தமிழ் எங்கள் உயிர்' என்ற பிரச்சார இயக்கத்தை நடத்தினார். இதன்வழி, 40,000 வெள்ளி ரொக்கமும் நூல்களும் திரட்டப்பட்டன.

1956 முதல் 1968 வரை தமிழகத்தைச் சார்ந்த அறிஞர்கள் தலைமைப் பொறுப்பை ஏற்றிருந்தனர். 1969 முதல் மண்ணின் மைந்தர்கள் தலைமைப் பொறுப்பை ஏற்று இத்துறையைத் திறம்பட வழிநடத்தினர். இந்திய ஆய்வியல் துறையில் கல்வியினை மேற்கொண்ட பலர் இன்று நல்ல நிலையில் உள்ளனர். தமிழ் சார்ந்த துறைகள் மட்டுமின்றி அரசாங்க அதிகாரிகளாகவும் தனியார்த் துறைகளிலும், வாணிபத் துறையிலும் பலர் சிறந்து விளங்குகின்றனர்.

இந்திய ஆய்வியல் துறையில் இதுவரை ஏறக்குறைய 20 பேர் முனைவர் பட்டமும் 50 பேர் முதுகலை பட்டமும் பெற்றுள்ளனர். இன்றைய நிலவரப்படி (ஏப்ரல் 2010) 40 பேர் முதுகலை பட்டத்திற்கும், 16 பேர் முனைவர் பட்டத்திற்கும் பல்வேறு துறைகளில் பட்டப்படிப்பை மேற்கொண்டு வருகின்றனர்.

மலாயாப் பல்கலைக்கழகக் கலை சமூக அறிவியல் புலத்தில் இந்திய ஆய்வியல் துறையில் இளங்கலை பட்டப் படிப்பிற்கான பாடங்கள்

ஆண்டு	பருவம் 1	பருவம் 2
ஆண்டு 1	தற்காலத் தமிழ் இலக்கியம் முற்கால இலக்கியம்	பயன்பாட்டு இலக்கியம் இடைக்கால இலக்கியம்
ஆண்டு 2	• சங்க இலக்கியம் • காப்பிய இலக்கியம் இலக்கியத் திறனாய்வு ஜோதிடக் கலை தமிழ்த்திரைப்பட வரலாறும் தாக்கமும்	• தமிழ் ஒலியனியல், யாப்பியல் வரலாறு • தற்காலத் தமிழ் இலக்கியம் - பாடநூல் நிதி இலக்கியம் நாடகக் கலை
ஆண்டு 3	• ஆய்வேடு தயாரித்தல் • தமிழ் ஒப்பு இலக்கியம் தமிழ் உருபனியல், தொடரியல் மலேசியத் தமிழ் இலக்கியம் தமிழ் நாட்டுப்புறவியல்	• சைவ , வைணவ பக்தி இலக்கியம் • ஆய்வேடு தயாரித்தல் தமிழ் பிரபந்த இலக்கியம் தமிழ் இலக்கியத் திறனாய்வு சிங்கப்பூர், ஸ்ரீலங்கா தமிழ் இலக்கிய வரலாறு தமிழ் பாரம்பரியத்தில் சித்தர் சிந்தனைகள்

• கட்டாயப் பாடம்

மலாயாப் பல்கலைக்கழக மொழி/ மொழியியல் புலம்

மலாயாப் பல்கலைக்கழகத்தின் மொழி மையத்தில் நீண்ட காலமாகத் தமிழ் இரண்டாம் மொழியாகக் கற்பிக்கப்பட்டு வந்தது. மொழி மையம் 1995-ஆம் ஆகஸ்டில் புலமாக உயர்வு கண்டது. 1998-ஆம் ஜூலை முதல் தமிழிலேயே இளங்கலை மொழியியல் பட்டப்படிப்பை மேற்கொள்வதற்கான திட்டம் அமல்படுத்தப்பட்டது. ஒவ்வொரு ஆண்டும் 15 முதல் 25 மாணவர்கள் வரை இளங்கலை மொழியியல் பட்டப்படிப்பை மேற்கொண்டு வருகின்றனர்.

மலாயாப் பல்கலைக்கழகத்தில் மொழி மொழியியல் புலத்தில் இளங்கலை மாணவர்களுக்குக் கற்பிக்கப்படும் தமிழ்ப்பாடங்கள்

<p>ஆண்டு 1 - பருவம் 1</p> <p>தமிழ் மொழித்திறன் 1</p> <p>மொழியும் இலக்கியமும் 1</p> <p>ஒலியனியல்</p> <p>ஆண்டு 1 - பருவம் 2</p> <p>தமிழ் மொழித்திறன் 2</p> <p>மொழியும் இலக்கியமும் 2</p> <p>உருபனியல்</p>	<p>ஆண்டு 2 - பருவம் 1</p> <p>தமிழ் மொழித்திறன் 3</p> <p>மொழியும் நாகரீகமும் 1</p> <p>தொடரியல்</p> <p>ஆண்டு 2 - பருவம் 2</p> <p>தமிழ் மொழித்திறன் 4</p> <p>சிறப்பு தமிழ்மொழி நோக்கத்திற்காகத்</p> <p>உரைக்கோவை</p>
--	--

<p>ஆண்டு 3 பருவம் 1</p> <p>தமிழ் மொழித்திறன் 5</p> <p>தமிழ்மொழி பரிணாம வளர்ச்சி</p> <p>இருவழி மொழிபெயர்ப்பு 1</p>	<p>ஆண்டு 3 பருவம் 2</p> <p>தமிழ் மொழித்திறன் 6</p> <p>இருவழி மொழிபெயர்ப்பு 2</p> <p>வட்டார வழக்கு</p> <p>ஆய்வு அறிக்கை</p>
---	--

குறிப்பு : மலாயாப் பல்கலைக்கழக, மொழி-மொழியியல் புலத்தின், மலேசியமொழிகள் மற்றும் பயனாக்க மொழியியல் துறையில், இந்திய மொழிகளின் நடுவண் நிறுவனத்தின் முன்னாள் துணை இயக்குநர், முனைவர் சாம் மோகன் லால் அவர்களும், தஞ்சை தமிழ்ப்பல்கலைக்கழகத்தின் முன்னாள் துணைவேந்தர் முனைவர் கி.கருணாகரன் அவர்களும் வருகைதரு பேராசிரியர்களாகப் பணியாற்றி வருகின்றனர் என்பது குறிப்பிடத்தக்கது.

மலேசிய திறந்தவெளிப் பல்கலைக்கழகம் (OUM)

மலேசிய திறந்தவெளிப் பல்கலைக்கழகம், கல்வி அமைச்சின் ஆசிரியர் கல்விப் பிரிவோடு இணைந்து நாடு முழுமையும் ஆசிரியர்கள் பட்டப்படிப்பை மேற்கொள்ள பயிற்சியினை நடத்தி வருகின்றது. 1996-ஆம் ஆண்டு, முதல் தமிழ் ஆசிரியர் பட்டப்படிப்பு தொடங்கப்பட்டது. நாடு முழுமையும் உள்ள ஆசிரியர் பயிற்சிக் கழகங்களில் சுமார் 300 தமிழ் ஆசிரியர்கள் வார இறுதிநாட்களில் பட்டப்படிப்பு மேற்கொண்டு வருகின்றனர்.

மலேசிய திறந்தவெளிப் பல்கலைக்கழகம் தமிழ்ப்பட்டப்படிப்புப் பயிற்சிக்காக இதுவரை 10 சிப்பங்களைத் தயாரித்துள்ளது. இவற்றில் 4 சிப்பங்கள் இலக்கணத்திற்கும், 4 சிப்பங்கள் இலக்கியத்திற்கும், 2 சிப்பங்கள் பயிற்றியலுக்கும் உரியன. அறிஞர்களைக் கொண்டு இச்சிப்பங்கள் தரமாகத் தயாரிக்கப்பட்டுள்ளன என்பது குறிப்பிடத்தக்கது.

இரண்டாம் மொழியாகத் தமிழ்

பின் குறிப்பிடப்பட்டுள்ள பல்கலைக்கழகங்களில் தமிழ் இரண்டாம் மொழியாகக் கற்பிக்கப்பட்டு வருகின்றது.

1. மாரா பல்கலைக்கழகம் (MARA)
2. மலேசிய புத்ரா பல்கலைக்கழகம் (UPM)
3. மலேசிய அறிவியல் பல்கலைக்கழகம் (USM)
4. மலேசிய சபா பல்கலைக்கழகம் (UMS)

சுல்தான் இட்ரீஸ் ஆசிரியர் பல்கலைக்கழகத்தில் (UPSI)

சுல்தான் இட்ரீஸ் ஆசிரியர் பல்கலைக்கழகத்தில் (UPSI) 2010-ஆம் ஜூலை தொடக்கம், இடைநிலைப்பள்ளித் தமிழ் ஆசிரியர்களுக்கான இளங்கலைப் பட்டப்படிப்புத் திட்டம் மேற்கொள்ளப்படவுள்ளது. அதற்கான பணிகள் தொடங்கப்பட்டு விட்டன. முனைவர் பட்டம்பெற்ற முழு நேர விரிவுரையாளரும், மூன்று தனிமுறைப் பயிற்றுநர்களும் தெரிவு செய்யப்பட்டுள்ளனர். இலக்கியம், இலக்கிய மரபு, இலக்கணம், மொழியியல், பயிற்றியல் ஆகிய ஐந்து துறைகள் அடையாளங் காணப்பட்டுள்ளன. முதற்கட்டமாக 2010ஆம் கல்வியாண்டில் 60 மாணவர்கள் தங்களது இளங்கலை பட்டப்படிப்பை மேற்கொள்ளவுள்ளனர்.

ஆசிரியர்க் கல்விக்கழகம்

நாடு முழுமையும் 28 ஆசிரியர்க் கல்விக்கழகங்கள் செயற்பட்டு வருகின்றன. அவற்றில் ஆறு ஆசிரியர்க் கல்விக்கழகங்கள் தமிழாசிரியர்க்கான பயிற்சியினை வழங்கி வருகின்றன. தமிழ் ஆசிரியர்க்கான முழுநேர பட்டப்படிப்பு, விடுமுறைக்காலப் பட்டப்படிப்பு, பட்டத்திற்குப் பிந்திய பயிற்சி, பணியிலுள்ள ஆசிரியர்க்கான பணியிடைப் பயிற்சி, குறும்பயிற்சி போன்ற பல விளைபயன்மிக்க பயிற்சிகளை ஆசிரியர்க் கல்விக்கழகங்கள் வழங்கி வருகின்றன.

நிறைவு

மலேசியாவில், மழலையர்ப் பள்ளி முதல் பல்கலைக்கழகம் வரையிலான தமிழ்க் கல்வியின் தொடர்ச்சியும் வளர்ச்சியும் மிகவும் சிறப்பான நிலையில் உள்ளது. மலேசியாவின் ஒன்பது மாநிலங்களில் தமிழ்ப்பள்ளிகள் உள்ளன. இம்மாநிலங்களில் தமிழ்ப்பள்ளி அமைப்பாளர்களும், தமிழ்மொழிக்கான துணை இயக்குநர்களும் தமிழ்க்கல்வி வளர்ச்சிக்காக அரும்பணி ஆற்றி வருகின்றனர். இதைத்தவிர கல்வி அமைச்சின் முக்கியப் பிரிவுகளான, கலைத்திட்ட மேம்பாட்டு மையம், பள்ளி ஆய்நர் பிரிவு, பாடப் புத்தகப் பிரிவு, ஆசிரியர்

கல்விப்பிரிவு, தேர்வு வாரியம் போன்ற பிரிவுகளிலும் தமிழ்க்கல்விக்கான துணை இயக்குநர்களும் அலுவலர்களும் திறம்பட செயலாற்றி வருகின்றனர்.

மேற்கோள்:

1. Arasaratnam, S. 1970. Indians in Malaysia and Singapore. London : Oxford University Press
2. Michael Stensen, p153,159; Class, Race & Colonialism in West Malaysia, 1980. Quoted from Indian Daily Mail 19.8.1946
3. Xavier S.Thani Nayagam, 1968, pp215,216, Tamil Studies Abroad a Symposium)
4. மன்னர் மன்னன். ம. 2009. உயர்க்கல்விக் கூடங்களில் தமிழ் : மெய்மையும் சவால்களும். பன்னாட்டுத் தமிழ் மொழியியல் மாநாடு. கோலாலம்பூர் : மலாயாப் பல்கலைக்கழகம்
5. சுப்பிரமணி. சோ. 2010. விடுதலைக்குப் பின் தமிழ்க்கல்வி : பரிணாமமும் சவால்களும். இந்திய ஆய்விதழ் துறை **ஆய்விதழ்**. கோலாலம்பூர் : மலாயாப் பல்கலைக்கழகம்

Moodle: A Tool for Tamil Teaching

K.Sarveswaran
Sri Lanka
iamsarves@gmail.com

Prof. V.Nagarajan
India
radius.consultancy@gmail.com

Abstract

Tamil language teaching has become an important need in the Tamil Diaspora, especially for the young generation. Constructing knowledge through interactions and collaboration is proved as an effect and efficiency method, which is referred as social constructivist learning, especially when the learners have some prior knowledge. The web 2.0 encourages collaboration. The e-Learning 2.0 is proposed based on Web 2.0 and constructivist learning. Moodle is a Free and Open Source Learning Management Software. The Moodle is developed based on E-Learning 2.0. As a result the Moodle provides space for collaborative learning through new cutting edge web 2.0 technologies, which in turn provides for effective learning. The young generation of Tamil society is familiar about Web 2.0. Therefore the Moodle is a good tool for Tamil language teaching.

1.0 Introduction

1.1 Social constructivist learning

Pedagogy refers to strategies of instruction, or a style of instruction [1] that is used to construct knowledge. Teachers use different styles of instructions to construct knowledge based on their experiences. A teacher can not convey knowledge to every student on the same level during a lesson, however implementing a variety of instruction styles in a course allows all the students to learn in at least one way that matches their learning style. Researches has been showing that the best way to learn is by having students construct their own knowledge instead of having someone construct it for them. The instruction style that help to construct knowledge in this way is called as Constructivist Learning[2]. The prior knowledge of the learners is important and it may help constructing knowledge. The prior knowledge comes from past experiences, culture, and their environment of the learners [2]. Therefore even when teacher constructs new knowledge the prior knowledge of the learners should be considered.

Social constructivist learning extends the constructivism into social setup and it is defined as groups of learners construct knowledge for one another, collaboratively creating a small culture of shared artifacts with shared meanings. When one is immersed within a culture like this, one is learning all the time about how to be a part of that culture, on many levels [3].

The knowledge can be shared in the forms of audio, video, text, 3D animations, question and answer, discussions, etc[2]. The percentages in Table 1 represents the average amount of information that is retained through different forms of knowledge [2].

Table 1 : average amount of information retained vs style of instruction

Style of instruction	Average amount of information retained
Lecture	5%
Reading	10%
Audiovisual	20%
Demonstration	30%
Discussion Group	50%
Practice by doing	75%
Teach others / immediate use of learning	90%

Though these forms can be used to some extent in face- face scenarios, when it comes to remote teaching it was difficult to share these forms of knowledge among learners in the past.

1.2 Web 2.0

Web 2.0 is the current standard of Web. The earlier web 1.0 was the read only web and the web site authors only could share the knowledge and the people who read could not share the reflections or their knowledge. Web 2.0 introduces many techniques that let authors and readers to share the knowledge. The techniques are Wiki, Forum, Blog, Chat, Web group. Social networking is also a resultant of all these techniques of Web 2.0.

1.3 E-Learning 2.0

e-Learning can be explained as learning that is supported through an electronic medium [4]. Due to the feature and facilities of the web is heavily used for e-Learning. The E-Learning 2.0 is came with the arrival of Web 2.0. The version 2.0 of electronic learning encourages the collaborative learning, constructing knowledge through Web 2.0 technologies such as Wiki, Forum, Mailing groups, social networking, blog etc.

1.4 Moodle

Moodle is an Open Source Learning Management Software, which is developed based on E-Learning 2.0 standard and the principles of social constructionist pedagogy. Moodle provides many activities related to collaborative learning such as Wikis, Blogs and Forum to implement this pedagogy. Moreover, the software is currently available in more than 70 languages including Tamil and also been used in more than 200 countries. All around the world there are many courses that are from K12, Undergraduate and Post graduate courses are conducted through Moodle. In Sri Lanka recently the University of Jaffna has started a Online BBM in Tamil medium using Moodle. However Moodle is been used in Sri Lanka for more than 5 years.

2.0 Teaching and Learning through Moodle

Moodle facilitates for the traditional teacher-led learning and the new type called social constructivist learning that is discussed in the section 1.1. Also the Moodle can be used as a part of blended learning, where the learning is conducted both face to face and e-Learning modes. The main components of Moodle are resources, Activities and other collaborative features like participants, messaging, calendar etc

2.1 Resources

According to the teacher led methodology the teacher can share their knowledge in form of Microsoft Word and Microsoft Powerpoint, Flash, Video formats, Audio formats and Portable Document formats (pdfs). More importantly these can be viewed in the Moodle it self without any other supporting tools. In terms of Moodle all these are referred as resources. In addition to these resources, it also lets to create Internal web pages (HTML formatted) using inbuilt HTML editor, Internal text pages and also facilitate to link Files that are stored locally or in remote locations such as web. Moodle also has a file repository to which we can upload files and can use them. All types of need of learners can be served by Moodle as it is supporting for many types of contents.

Not only just having resources but also those can be arranged according to the instruction style a teacher would like to follow. The resources can be easily added and edited. Figure 1 shows a part of Moodle page where some resources are used.

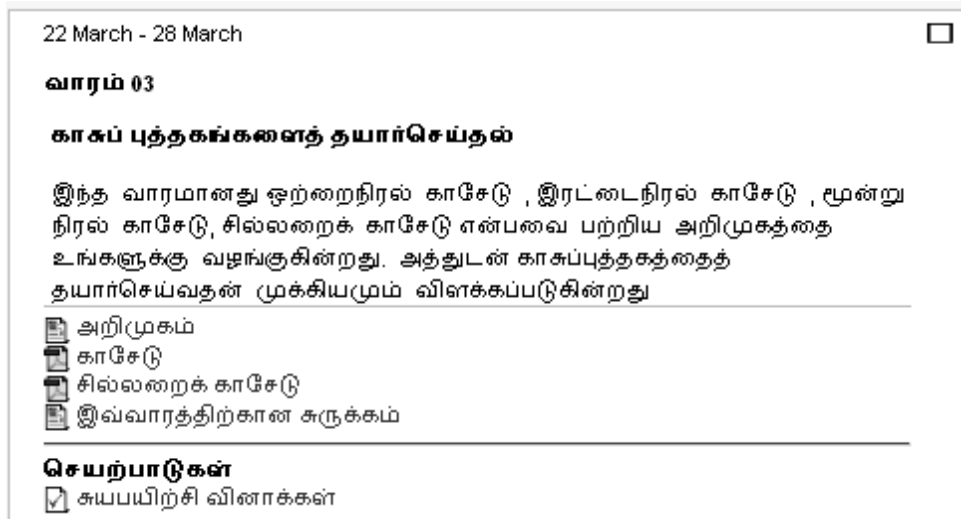


Figure 1 : Resources view in Moodle

Like wise in social constructivist learning point of view also these resources can be used to share knowledge of learners and help constructing knowledge.

2.2 Activities

In addition to the resources, activities are another main part of Moodle. Assignments, Wiki, Choices, Forum, Lessons, Quizzes and Surveys are some of the activities that are used in Moodle. In teacher led learning these activities are used to assess learners.

On the other hand these activities are used as a medium to construct knowledge in social constructivist learning. The learners can collaboratively participate in Forums, Wiki, Blogs etc and share their experiences and knowledge with peers to construct knowledge. The facilitators can monitor activities and administrate activities.

2.3 Other Collaborative features

Moodle has chat and email to interact among peers and teachers. In addition to that anyone can maintain a blog and update their profiles so that they can get to know each other. Also there is a calendar to manage the events and in which we can keep global events and personal events.

2.4 Reporting and Monitoring

Moodle has very rich monitoring mechanism through which every single action of users can be monitored. We can track from which IP the user is logged in and what are the pages viewed and what are the actions performed during the visit. In addition that Moodle shows the recent events when a users log in and that help one to check what has happened recently.

2.5 Administration and Security

User administration, Course administration and System administration can be easily done in Moodle. Under user administration we can create and administrate users as well as we can control the enrollments. Since Moodle is a web based system the security considerations have been taken well. A usage policy, user privileges customization, automatic backups, secure logins etc can be managed and maintained.

3.0 Methodology

As Moodle is a Learning Management System, users can only interact with courses. Therefore before start any lessons, a course need to be created.

Next the teachers who are responsible for the course should be assigned. Teachers have over all control of the courses and they can perform all the actions as they like. In terms of e-Learning they are called as facilitators, not teachers. The actual responsibility of the facilitator is to make the course and facilitates learners so that the learners can easily interact with each other and construct knowledge.

After that the relevant learners should be assigned to the course. Users even can be further grouped inside a course so that they can be managed easily.

Next create relevant activities so that learners can interact with each other and share knowledge. The activities should be created so that when learners can follow them and the intended learning objectives and knowledge can be achieved. In addition to that the facilitators should monitor and check whether information overloading and unnecessary knowledge sharing are happening.

4.0 Recent developments

Virtual worlds may useful in future to conduct and practical oriented courses via online. Soodle [7] is an effort to have the Moodle installed in Virtual worlds and connect it with real world Moodle instants. Nowadays people have started buy lands in virtual worlds and many vendors have even started their online stalls in virtual worlds. Moodle also has stepped in to that and many researches going in this directions.

Further many the activities, like role play, are still not possible in Moodle. The next major release of Moodle is Moodle 2.0 and it has many improvements including more support for collaboration, security, web services, repository managements etc.

5.0 Conclusion

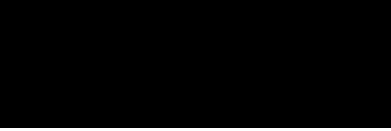
Moodle is a good tool for collaborative learning. The young Tamil generation is more interested in social networking and virtual collaborations. Therefore the collaborative learning which is facilitated by Moodle Learning Management System may help them to learn Tamil language efficiently and effectively.

6.0 References :

6. <http://en.wikipedia.org/wiki/Pedagogy>, accessed on 20100420.
7. Brooks, J. and Brooks, M. (1993). *In Search of Understanding: The Case for Constructivist Classrooms*, ASCD
8. <http://docs.moodle.org/en/Philosophy>, accessed on 20100320.
9. <http://en.wikipedia.org/wiki/Elearning>, accessed on 20100420.
10. <http://uoj.nodes.lk>, accessed on 20100520.
11. <http://demo.moodle.net/course/view.php?id=615>, accessed on 20100520.
12. <http://www.sloodle.org/moodle/>, accessed on 20100520.

3

கணினி மொழியியல்



Morphological Generator for Tamil

A new data driven approach

Rekha R U, Anand kumar M, Dhanalakshmi V Soman K P

Computational Engineering and Networking
Amrita Vishwa Vidyapeetham Coimbatore,
{m_anandkumar,v_dhanalakshmi,kp_soman}@cb.amrita.edu

Rajendran S

Department of Linguistics
Tamil University, Thanjavur, India
raj_ushush@yahoo.com

Abstract

Tamil is morphologically rich language. Being agglutinative language most of the categories expressed are suffixes. Tamil is a post positional inflectional language it has more suffixes compared with English. The Morphological Generator takes lemma and a morpho-lexical description as input and gives a word-form as output. It is a reverse process of Morphological Analyzer. Morphological generator system implemented here is a new data driven approach which is simple, efficient and does not require any rules and morpheme dictionary. We have developed an individual system to handle nouns and verbs. Any automated machine translation system requires morphological analyzer of source language and morphological generator of the target language. Using this morphological generator we also developed a verb conjugations and noun declension. Here Tamil verbs are classified into 32 paradigms [1] and 1500 word forms are handled. Like verbs nouns are classified into 25 paradigms and 325 word forms are handled. An inflection table is arranged in Two Dimensional format where row corresponds to the morpho lexical form and column corresponds to the paradigm number. The noun inflection table contains 325 rows (word forms) and 25 columns (paradigms) similarly verb inflection table contains 1500 rows (word forms) and 25 columns (paradigms).

1)Introduction

Natural Language Processing (NLP) has been developed in 1960. The aim of NLP is studying the problems in the automatic generation and understanding of natural languages. The primary goal is to build computational models of natural language for its various analysis and generation. Tamil verbs are inflected into several grammatical features. In Tamil language the verb specifies almost everything like gender, number, and person markings and also with auxiliaries it represents mood and aspect. These are the morphological information of the root words. This makes a challenging work in Tamil. In general in Indian language there are many inflections compared to other languages. Morphological generator generates a word form from a lemma, word class, and the type of *morpho-lexical inflection* required. In Tamil language some time the root word undergoes morphological change when it attaches to the inflection. Morphological generator can be an individual module or integrated with several NLP applications like machine translation, Automatic sentence generation. In this paper we describe a fast and

simple morphological generator for Tamil. This novel approach can be applied to any morphologically rich languages.

2) Morphological Generator for Tamil

Generally, morphological generator tool is developed using rule based approach. Where the rule based approach requires a set of morphosyntactic rules, spelling rules and morpheme dictionary. In this novel approach rules and dictionaries are not required it only requires the inflection table and paradigm classifier program. Here ,the morphological generator receives an input in the form of *lemma+word_class+morpho-lexical Information*, where *lemma* specifies the lemma of the word form to be generated, *word_class* specifies the grammatical category (noun/verb) and *morpho-lexical Information* specifies the type of inflection.

In this section we describe the files used in the data creation and the algorithm for the implementation of this system. A Perl program is written for finding the paradigm and index number it makes this system simple and efficient. The algorithm of this method is described in the below sections

2.1) Creation of Inflection Table

Number of paradigms for each word class (noun/verb) is defined .In Tamil; there are 32 paradigms for verb and 25 for noun. For every paradigm a word is selected this is termed as head word. For this head word all word form is created, in Tamil there are more than thousand word forms are possible for a head word. Here we have selected 1500 most frequently used word forms for verb including auxiliary& clitics and for noun it is 325 including postpositions. This verb/noun word form creation uses an order which is followed for all the paradigms. A *morpho-lexical Information* list is also created for the above word forms. Using all the word forms a table is created each column of the table corresponds to its paradigm. For each column remove the stem. This table is converted into a tabular CSV format and represented as an *inflection table*.Table.1 represents the sample data for Tamil verbs. In this table row indicates the Morpho-lexical inflection and column indicates paradigm number.

		Paradigm number						
		1	2	3	4	5	32
<i>Morpho-lexical Inflection</i>	1	<i>ththAn</i>	<i>inAn</i>	<i>NdAn</i>	<i>ddAn</i>	<i>RAn</i>		
	2	<i>ththAL</i>	<i>inAL</i>	<i>NdAL</i>	<i>ddAL</i>	<i>RAL</i>		
	3	<i>ththAR</i>	<i>inAR</i>	<i>NdAR</i>	<i>ddAR</i>	<i>RAR</i>		
	4	<i>ththOm</i>	<i>inOm</i>	<i>NdOm</i>	<i>ddOm</i>	<i>ROm</i>		
	:							
:								
	1660							

Table 1 . *Inflection table for Tamil Verbs*

Input Format :

Lemma + word class + morpho-lexical Information

Example:

padi +V+ PAST_3SM = padiththAn.

maram+N+ ACC_Marker = maraththai.

2.2) Algorithm Developed for Morphological Generator

Input Format : Lemma + word class + morpho-lexical Information

Out format : *Word*=MORPGEN(*Input*)

MORPGEN(*Input*)

l,w,m =SPLIT(*Input*)

rl=ROMAN(*l*)

parnum=PARNUM(*rl,w*)

colindex=*parnum*

rowindex=INDEX(*m,w*)

inflection=INFTABLE[*rowindex*][*colindex*]

stem=STEM(*l,w,parnum*)

word=JOIN(*stem,inflection*)

end

where ,

l is the lemma

w is the word class

m is the morpho-lexical information

rl is the romanized lemma

SPLIT: This function split the users input into as lemma, word class and morpho-lexical information.

ROMAN:It romanizes the lemma part.

PARNUM :This function identifies the paradigm number using romanized lemma and word class

INDEX :This retrieves the index of morpho-lexical inflection from the morpho-lexical inflection

file

INFTABLE: Using the row and column index the inflection part is retrieved from the inflection

table

STEM : Taking romanized lemma and paradigm number as input it gives the stem

JOIN : It concatenates the stem and inflection part

3) Methodology

The whole system is divided into two modules. Module-I handles the lemma/root part and Module-II handles the morpho-lexical information. In Module-I the lemma/root word in Unicode format is romanized and the paradigm number is identified by end characters. A simple PERL program is written for finding the paradigm number. The identified paradigm number is referred as *Column Index* and stemming is also performed using this paradigm number. In Module -II the Morpho-lexicon information is given as an input. A complete set of Morpho-lexicon information are stored in a file with index numbers.

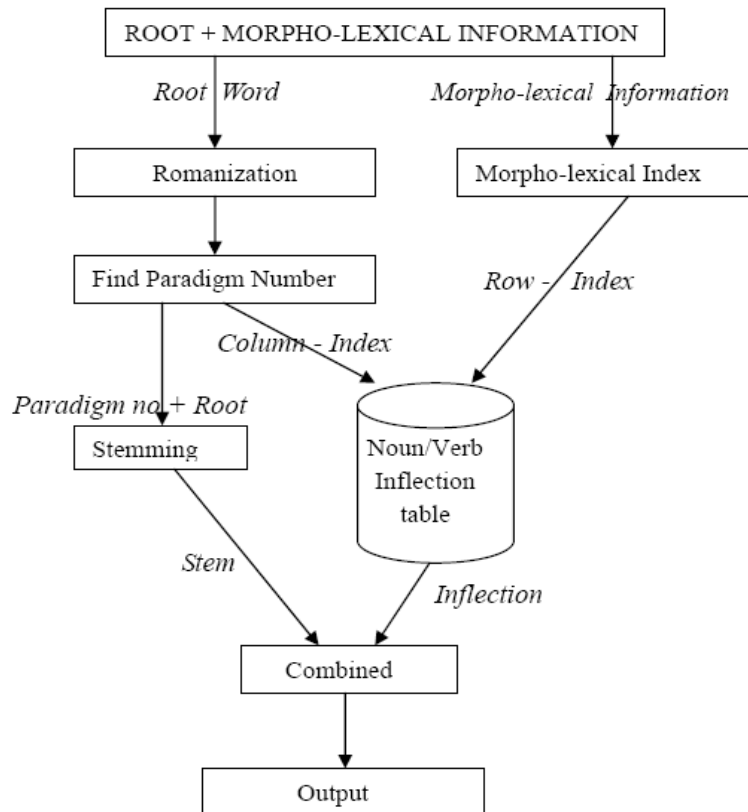


Figure 1: Morphological Generator System

The index number of the corresponding input is identified. This refers the *Row Index*. A verb and noun inflection tables are used in this system. In this Two-Dimensional inflection table rows are Morpho lexical information index and columns are paradigm numbers. For each paradigm we have created a complete set of morphological inflections corresponding to the morpho lexical information. Finally using the *column index* and *row index* morphological inflection is retrieved from the inflection table. This inflected form is affixed with the stem. In this work a morphological generator is designed for each of the syntactic categories and then combined to morphologically generate a complete sentence.

In this system we also handled some difficult tasks. First one is where a lemma has more than one possible surface word form given a particular morpho-lexical inflection type and word class. Second is where a surface word has more than one possible morpho-lexical inflection type. The gui snapshot for the morphological generator is shown in figure 2.

Examples:

படி + V + PAST_3SH = படித்தார்/படிந்தார்

ஓடு + V + PAST_3SF = ஓடினாள்

சா(சு)+V+ PAST_3SM =செத்தான்

மரம்+N+PL=மரங்கள்

முள்+N+ACC=முள்ளை

பூ+N+PL_DAT=பூக்களிற்கு

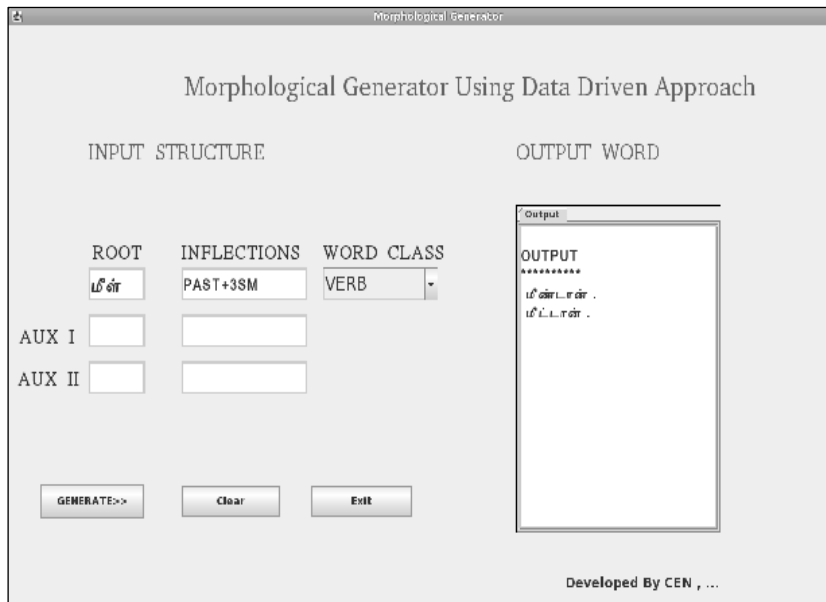


Figure 2 : GUI of morphological generator

Conclusion and Future work

Morphological generator which is explained here is a novel approach .It is developed using a very simple and efficient method. This is not a language specific method .so this can be applicable for all the morphologically rich languages. Using this approach currently we are developing morphological generator for Malayalam and Telugu. This system provides a vast application it's used in noun declension, verb conjugation, automatic sentence generator and also in featured .this system is unique handles auxiliaries and clitics, it does not require any spelling rules and this methodology can be implemented for any language .This work can be further used for implementing morphology based translation system between any language to Tamil.

References

1. Rajendran, S., Arulmozi, S., Ramesh Kumar, Viswanathan, S. 2001. 'Computational morphology of verbal complex . languageinindia Volume 3 : 4 April 2003
2. Guido Minnen, John Carroll, and Darren Pearce. 2000. Robust applied morphological generation. In Proceedings of the First International Natural Language Generation Conference, pages 201.208, 12.16 June.
3. Ganapathiraju, Madhavi and Lori Levin: TelMore: Morphological Generator for Telugu Nouns and verbs, In the proceedings of Second International Conference on Universal Digital Library Alexandria, Egypt November 17--19, 2006
4. Anandan, P, Rajani Parthasarathy, Geetha, T.V. 2001. "Morphological Generator for Tamil" in Tamil Internet 2001 Conference Proceedings, Malaysia.

Examining verbal forms inside the Tēvāram, in the light of the vocabulary found inside the 9th chapter of Cēntaṅ Tivākaram

Dr Jean-Luc Chevillard
CNRS, University Paris-Diderot
Paris7, UMR7597 (HTL)
jean-luc.chevillard@univ-paris-diderot.fr

Introduction

The goal of the present communication is to examine a feature of a possible vocabulary data base, which could make more precise our knowledge of the history of Tamil vocabulary across the centuries, and especially that part of the vocabulary which consists of verbs. The intended final target of the article is to present information structured in such a way that it would be possible to make a comparison between two texts:

one of them, the *Tivākaram*, is the most ancient Tamil Thesaurus (or *kōśa*) and belongs to technical/scholarly literature

the other one, the *Ēḷān Tirumuṭai*, alias Cuntarar's *Tēvāram*, belongs to devotional literature.¹

Due to space and time limitation, an effective comparison remains of course a distant goal. Moreover, I start by providing the reader with a relatively detailed presentation of the *Tivākaram*, a text probably not known well enough inside and outside Tamil Nadu, before examining the presence (or the absence) of the vocabulary of the *Tivākaram* inside Cuntarar's *Tēvāram* and comparing this with what is found in other texts.

Structure and function of the *Tivākaram*

The *Tivākaram*, (also called *Cēntaṅ Tivākaram*), is an object probably not very familiar to modern users of Tamil, unless they have received special training from a traditional Tamil scholar.² However, some of them, having lived in the English-speaking world might in fact be familiar with some version of a standard reference work that has a lot in common with the *Tivākaram*, namely the "wordfinder" known

1 As regards the second one, there was no special reason for selecting it, except that it was available to me in digital form (See Subramanya Ayyar at alii [2007]) and that the two texts are probably not very distant in time.

2 Another possibility might be their having read some of the articles written by G. James, like for instance James[1989].

as *Roget's Thesaurus*.³ The resemblance can be hinted at in the following way: whereas we can see inside *Roget's Thesaurus* an entry such as:

- (1) #360. **Death.** -- N. death; decease, demise; dissolution, departure,[...] V. die, expire, perish; meet one's death, meet one's end; pass away, be taken; yield one's breath, resign one's breath; resign one's being, [...]⁴

we find, similarly, inside the *Tivākaram*, an entry which reads:

- (2) T1674 *tuñcal, poṇṇal, viṭtal, vītal, muñcal, tuvaṇṇal, muṭṭovu, ulattal, vīṭal, māyṭal, iṇṭtal, māḷal, tapuṭal, cātal eṇṇac cāṇṇar pulavar.*

Modern editors have added intermediate titles (or labels) to the *Tivākaram* and this *cūttiram* is labelled by them “ (*cātalīṇ peyar*) ”, which seems to make it clear that in their view, T1674⁵ is an enumerations of 13 quasi-synonyms to *cātal* “ to die ”. I use the expression “ quasi-synonyms ” because just as English speakers will recognize that “ to expire ” or “ to perish ”, appearing in 1, are not exact synonyms of “ to die ”, similarly the users of Classical Tamil know that the 13 terms enumerated in 2 are not exact equivalents of *cātal*, although the “ nuances ” are not easy to render into English.⁶

Additionally, it must be added that although this article concentrates on verbal forms, the *Tivākaram* is basically an enumeration of nouns, many of which are proper names whereas others are verbal nouns, common nouns, etc. And it seems clear that the practical synonymy problem is not the same when it concerns the names of a God,⁷ the names of botanical species,⁸ and when it has to do with what the title of the 9th chapter in *Tivākaram* describes as *ceyal paṇṇiya peyart tokuti*, our main target here.

Organization of the 9th chapter in *Tivākaram*

I shall now try to give a concise characterization of that 9th chapter of *Tivākaram* (henceforth 9T), as it appears in the 1993 critical edition (Madras University). The main facts concerning it which must be mentioned in a preliminary abbreviated description are the following:

3 The most complet presentation of *Roget's Thesaurus* which I know of is Hüllen [2004], *A History of Roget's Thesaurus*. It must also be noted that Knuth's 1993 *Stanford GraphBase* contains a section devoted to *Roget's Thesaurus*.

4 This small citation is excerpted from the file available on the Project Gutenberg server (Edition 15a), itself based on the 1911 edition.

5 This reference, like every other reference to the *Tivākaram* in this article, is to the 1990-1993 two-volume critical edition by Mu. Caṇmukam Piḷḷai and I. Cuntaramūrtti.

6 The primary meaning of the first item (*tuñcal*) is “to fall asleep”; but readers of TC17c know that *tuñciṇṇar* “they fell asleep” is a well-known euphemism (*takuti*) for *cettār* “they died” (see Chevillard [1996:65]).

7 The *Tivākaram* is divided into 12 chapters and the first chapter gives the names of Gods. For instance, *cūttiram* 2 provides us with a list of 64 names for Civaṇ, which starts thus: “*caṇṇakaraṇ, iṇṇiyōṇ, kāma takaṇṇaṇ, kaṇṇā taraṇṇē, [...]*”.

8 The 4th chapter of *Tivākaram* is called *Marappeyart Tokuti*. As an example, its 78th *cūttiram* (alias T705) provides us with a list of 5 alternate names for the plant generally called *tālai*: “*kaṇṇal, kēṭakai, maṭṭi, kaitai, muṭṭaṇkal, vaṇṇu imir tālai eṇa vakut taṇṇarē*”.

- 1)9T contains 213 cūttiram-s (numbered from T1558 to T1770 in the critical edition)⁹ and many of those entries belong to the type illustrated in 2.¹⁰ They are accompanied by a concluding verse (T1771) which is a praise of Cēntaṅ.
- 2)In the modern editions and in the first (1840) printed edition, each of the 213 cūttiram-s is preceded by a title, which has the form “X-iṅ peyar”, where X is generally a verbal noun or a verbal complex expression.¹¹
- 3)Item X, called by me in this article “primary entry”, is found again in the *cūttiram* itself, where it is accompanied by other items, which I shall call “secondary entries”. For instance, in T1674 (see 2), the primary entry is *cātal* and the secondary entries are the 13 items starting with *tuñcal*, *poṅṅal*, *viñtal*, etc. In what follows, locations of primary (resp. secondary) entries will be marked by “P” (resp. “S”).¹²
- 4)The relationship between a primary entry and an associated secondary entry seems to be that the latter can be used (at least in certain contexts) as an approximate synonym of the former.¹³
- 5)The total number of secondary entries associated with the 213 primary entries found in 9T is 673, which gives us the approximate figure that each primary entry is associated, on the average, with a little more than three secondary entries.
- 6)It must be noted that in the count of 673, some items are counted several times, notably because the same secondary item can be associated with several distinct primary items. More details will be given later on this feature.

Since this article is concerned with Tamil verbs in Classical Tamil, I do not take into account in this study the items which are not lexemes classifiable as verbs.¹⁴ I shall only consider, in this preliminary exploration of the verbal vocabulary contained inside the *Tivākaram*, the primary and secondary items in 9T which end with a few clearly recognizable verbal suffixes, namely:

- 7)197 items ending in *-tal* or variants of *-tal*: *akattiñtal* [T1632], *akaṅal* [T1585], *acaital* [T1612, T1736]), ...
- 8)157 items ending in *-ttal*: *akaittal* [T1575, T1672], *acaittal* [T1682], ...
- 9)199 items ending in *-al*: *aṅṅai kuvittuk koṅṅal* [T1765], *aṅṅal* [T1672], *aṅṅaiṅkal* [T1675], *aṅṅal* [T1607],...

9 The 9th chapter in the 1840 uncritical edition (reproduced in the Santhi Sadhana Trust 2004 edition) contains 283 entries (followed by a concluding verse).

10 Exceptions are entries like 9T-1 (=T1558 “toḷiṅ peyar”), 9T-2 (=T1559 “toḷiṅ payilviṅ peyar”),

11 I have not verified whether those “titles” are found in manuscripts.

12 This means that the location of *cātal* (resp. *poṅṅal*) will be given as PT1674 (resp. MT1674).

13 I shall not try to demonstrate this fact and take as a working hypothesis that this is the “raison d’être” of a book like the *Tivākaram* (N.B. this assertion does not concern chapters 11 and 12).

14 That means that I shall not include in the preliminary data base primary items such as *kōpam* (T1581) “anger” or *tuṅṅai* (T1762), which is the name of a dance.

In the remaining part of this article, I shall concentrate on that subset of 553 items,¹⁵ examining the way it has to be normalized, in order to become usable for a confrontation with a textual corpus such as the *Tēvāram*.

Normalization of the 9th chapter in Tivākaram

As already indicated, some items are found in several *cūttiram*-s of 9T. I include in this “multiple occurrence” phenomenon the cases where the same lexeme occurs, for metrical reasons, under distinct but equivalent forms such as *aṭal* and *aṭutal*.¹⁶ Additionally, it must be noted that a few *cūttiram*-s in 9T are preferably understood as explanatory definitions: such is the case for instance for T1762, in which *tuṇarīkai* is explained or T1766 which provides another designation for the action of “*ārttu vāy puṭaittal*”.¹⁷ If we set those cases aside, we are left with a core of 442 Tamil verbs, among which:

- 102 items¹⁸ occur in exactly 1 *cūttiram*, where they are the primary entry: they will be said to have profile “P”
- 2 items¹⁹ occur in exactly 2 *cūttiram*-s as the primary entry: they have profile “PP”.
- 260 items²⁰ occur in exactly 1 *cūttiram*, as a secondary entry: they have profile “S”.
- 50 items²¹ occur in exactly 2 *cūttiram*-s, as a secondary entry: they have profile “SS”

More complicated profiles are also found, because some items appear both as primary entry in one *cūttiram* and as secondary entry in another one. The exact figures are as follows:²²

P	PP	S	SS	SSS	PS	SP	PPS	SPP	SSP	Total
102	2	260	50	5	8	12	1	1	1	442

Chart 1: Profiles of Verbal items in the 9th chapter of Tivākaram (9T)

15 Ulterior versions on this exploration should of course take into account other verbal items such as: (1) 24 items ending in *-vu*: {*āyvu* [T1619], *icāivu* [T1710], *kataḷvu* [T1587], ...} ; (2) 8 items ending in *-ppu*: {*iyaiṅṅu* [T1710], *uḷṅṅu* [T1615], ...} ; (3) 7 items ending in *-ci*: {*cārci* [T1722], *cūcci* [T1618], ...}.

16 Variants which I have noted are: *aṭal/aṭutal*, *aṇṇal/aṇṇutal*, *ukaṭal/ukaḷal*, *uṭṇal/uṭṇutal*, *eṇṇal/eṇṇal*, *oṇṇal/oṇṇutal*, *kaṇṇal/kaṇṇutal*, *kuṇṇal/kuṇṇutal*, *kūṭal/kūṭutal*, *ceyal/ceyṭal*, *tāḷal/tāḷutal*, *tōṇṇal/tōṇṇal*, *naṇṇal/naṇṇutal*, *nīṇṇal/nīṇṇutal*, *neruṇṇutal/neruṇṇal*, *nēṇṇal/nēṇṇal*, *viṭṭal/viṭal*.

17 T1762 reads: “*muṭṭakkai irukai paḷṅṅuṭai oṇṇit tuṭṭakkiya naṭṭiyatu tuṇarīkai ākum*” and T1766 reads: “*āvalarī koṭṭal, ārttuvāy puṭaittal*”.

18 This list of 102 items starts with: *akattiṭṭal* (PT1632), *aṭṭal* (PT1669), *aṇṇal* (PT1607), *avittal* (PT1744), *aḷṭṭal* (PT1652), *aḷṭal* (PT1678), *aḷṭal* (PT1580), *aḷṭal* (PT1650), *aṇṇutal* (PT1645), *ārāyṭal* (PT1619), *irattal* (PT1635), ...

19 Those two items are: *kalattal* (PT1643 & PT1723) and *neruṇṇutal/neruṇṇal* (PT1591 & PT1671).

20 This list of 260 items starts with: *akaṇṇal* (ST1585), *acaittal* (ST1682), *aṇṇarīkal* (ST1675), *aṇṇavaral* (ST1737), *aṇṇavutal* (ST1722), *aṇṇal* (ST1567), *anartal* (ST1735), *amaṇṇal* (ST1591), *amaṇṇal* (ST1733), *ayartal* (ST1560), *ayaravuyirttal* (ST1695), *arital* (ST1645), *aruntutal* (ST1570), *aruḷal* (ST1627), *alkutal* (ST1613), *alarīkarittal* (ST1567), *alartal* (ST1731), *avaittal* (ST1728), *aḷṭal* (ST1723), ...

21 This list of 50 items starts with: *akaittal* (ST1575 & ST1672), *aṭal/aṭutal* (ST1672 & ST1675), *aṇṇal/aṇṇutal* (ST1607 & ST1722), *ayital* (ST1569 & ST1570), *aḷṭal* (ST1627 & ST1692), *aṭal* (ST1686 & ST1752), *ārttal* (ST1682 & ST1686), *ārtal* (ST1570 & ST1612).

22 Due to space limitations, the detailed content of each category will be made available in an appendix, contained in an online publication.

The *Tivākaram* seen as a semantic graph

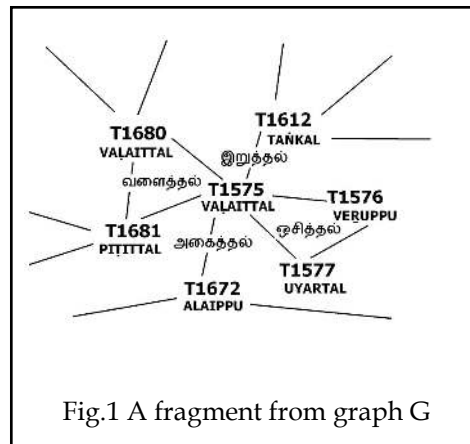
Having thus provided some elementary statistics on the verbal forms which are found inside the 9th chapter of the *Tivākaram*,²³ I shall now try to provide another perspective, concentrating on the *cūttiram*-s in which the verbal forms occur. As already said, the co-occurrence of two verbal nouns in the same *cūttiram* indicates an approximate synonymy: there are cases when the two verbs can be used to convey the same notional kernel, which is what the *cūttiram* itself shall be said here to represent. On the other hand, when a verbal noun is found in several *cūttiram*-s, it means that the verb is polysemic, and can be used (in different contexts) for expressing several different notions. This complexity can be represented by a graph G in the following way:

- each *cūttiram* of 9T in which there occurs one of the verbal nouns which we examine is said to be a **vertex** of graph G
- each time two *cūttiram*-s possess a verbal noun in common, we say that an **edge** of graph G connects the two corresponding vertices.

According to my calculations, the number of vertices to be considered here is 156 and the number of edges is 98. Additional statistics are as follows:

- The graph G contains 84 components, but 70 of those components contain only one element.
- Among the small components, 9 contain 2 elements; 2 have 3 elements and 2 have 4 elements
- There is a huge component containing 54 elements, and inside the huge component, seven 3-cliques are visible.

I have prepared a planar representation the graph G, but it is of course, due to space limitation, not possible to include it here. I shall content myself here with providing a small extract from the huge component, considering the fragment which contains the vertex T1575 (VAḶAITTAL) and the vertices which are connected to it through edges.



23 Since this fact has not yet been stated, it should be said that verbal lexemes are also found in other chapters of the *Tivākaram*. See e.g. T1799 (*aḷittaliṅ peyar*). I have not however tried to count them.

In this fragment, the vertex T1575 represents the notion of “ BENDING ”²⁴ which seems to be the common denominator, from a semantic point of view, between the various verbal nouns appearing inside *cūttiram* T1575, namely:

- (3) *ocittal, kōṭṭal, oṭṭal, akaittal,*
ṅemirtal, iṅuttal, vaḷaittal ākum (T1575)

Additionally, the element அகைத்தல் which appears on top of the edge connecting T1575 and T1672 (ALAIPPU) corresponds to the fact that *cūttiram* T1672 also contains the same element.

- (4) *ulaittal, oṅuttal, tirttal, aṭal, miṅaiceyal*
akaittal, vētarṅai, alaippu eṅa moḥpa (T1672)

An elementary test and some questions

After those preliminary explanations, we are now ready to present a few elements in order to give an idea of what could be attained by a systematic exploration of a corpus like the *Tēvāram* on the basis of the semantic graph which can be perceived inside the *Tivākaram*. Due to space limitation, and because of the huge nature of the task, I shall concentrate on one example, making use of the item examined above, namely VAḶAITTAL (T1575). The basic idea of each elementary test is to verify whether all of the approximate synonyms are equally represented inside a given corpus of text. In the case of Cuntarar’s *Tēvāram*, on the basis of T1575 (with 1 primary entry and 6 secondary entries), the results are provided in column 2 of the following chart, where they can be compared with the results of the same test on the other parts of *Tēvāram* (attributed to Campantar and to Appar) and with the *Puṅṅāṅṅūṅṅū* as another element for comparison:

Entry	Cuntarar ²⁵	Campantar	Appar	Puṅṅam ²⁶
<i>ocittal</i>	NO	1-136_2: <i>ocintu</i>	4-39_9: <i>ocitta</i>	ociya “ having broken ” (80-8)
<i>kōṭṭal</i>	NO	?? 1-128 (kōṭṭiṅār)	4-10_(4): <i>kōṭṭiṅār</i>	cf. kōṭṭaṭu (55-11)
<i>oṭṭal</i>	7-9_(10): <i>oṭṭikkum</i> “ it will destroy ”	[2-24_(8): <i>oṭṭintu</i> “ having been broken ”]	NO	“to break ”
<i>akaittal</i>	NO	NO	4-54_(1): <i>akaittiṭṭu</i>	NO
<i>ṅemirtal</i>	NO	NO	NO	NO
<i>iṅuttal</i>	7-2_10: <i>iṅuttir</i>	1-133_(7): <i>iṅuttavaṅ</i>	4-16_(10): <i>iṅuttār</i>	“to break ” (373-20)
<i>vaḷaittal</i>	7-57_(5): <i>vaḷaitta</i>	3-106_(7): <i>vaḷaittu</i>	6-42_(5): <i>vaḷaittu</i>	158-1 (rare)

Chart 2: one elementary test (T1575, VAḶAITTAL “ BENDING (or breaking?) ”)

24 The other VAḶAITTAL (T1680) which appears on the graph is “ SURROUNDING ”.

25 The occurrences for the 3 *Tēvāram* authors were found through the concordance contained in the Digital *Tēvāram* CD-ROM (see bibliography).

26 The examples are based on V.I. Subramoniam [1962].

As should be clear from this chart, the result of this single test is of course highly inconclusive, but it is nevertheless an occasion for asking difficult and unavoidable questions. Based on more tests, which I cannot show here because of space constraints, my impression is that rare (or archaic?) synonyms are found more often in Campantar and Appar's *Tēvāram*-s than in Cuntarar's. That may be due to the fact that we have only 100 hymns by Cuntarar whereas we are having 385 hymns by Campantar and 312 by Appar, and that some rare words appear only when a corpus is large enough. This might however also be an indication of changes in the lexicon.

Additional questions concerns the meaning: How can we know what was the meaning intended for each word by the person who composed the *Tivākaram*, especially since it did not reach us with a commentary? And would those original intended meanings be relevant here, if we knew them for sure? In particular here, since some of the examples quoted here concern an action of bending to the point of breaking or crushing and since others concern an action of bending without breaking,²⁷ my working hypothesis that all elements in T1575 (see 3) are approximate synonyms may in this case appear as partly problematic, but which other explanation could be given?

These questions being asked and being the basis for further exploration, I would like in conclusion to state that a parallel exploration of traditional lexicographical tools and literature can also contribute to a better understanding of the way a thesaurus like the *Tivākaram* was made (on the base of which compilation was it prepared?) and of the reasons why the need was felt later to create new tools, such as for instance the *Piṅkalam*, its immediate successor in time. The answer to such questions can only come from a joint examination of many pairs (Thesaurus/literary work), like the one (*Tivākaram*/*Tēvāram*) which has been sketched here. And I intend to continue my explorations in that direction. Finally, I also would like to point to the fact that the influence which traditional works such as those have had on literary composition is probably underestimated in the modern world. The following quotation, excerpted from the preface to a 1968 edition of *Piṅkalam*, gives food for thought.

“இலக்கண இலக்கிய நூல்களைக் கற்றுப் புலமைத் திறமடைய விரும்புவோர் முதற்கண் நிகண்டு நூல்களை ஐயந்திரிபறக் கற்று நல்ல பயிற்சியைப் பெற்றிருக்க வேண்டும் என்னும் கட்டாயத் திட்டம் பண்டைக்காலக் கணக்காயர்களிடத்திலே இருந்து வந்தது. இத் திட்டம் அண்மைக் காலம் வரையில் நடைமுறையில் இருந்தது. பழங்காலக் கணக்காயர்கள் நிகண்டு கற்று இலக்கியப் பயிற்சி பெற்ற மாணவர்கட்குத்தான் இலக்கணம் போதிப்பதென்னுங் கொள்கையைச் சிறிதும் நழுவவிடவில்லை. தமிழ் நாட்டிலே புதிய முறையில் பல்கலைக் கழகங்களும் கல்விச் சாலைகளும் பல்கிய பின்னர் பழங்கால முறை அடியோடு அடிபட்டுப் போய்விட்டது. முற்காலத்திற் றமிழ்மக்களுக்கு அகராதிகளாகப் பயன்பட்டவை நிகண்டுகளேயாம். மேலும் இந்நாள் அகராதி போன்று அந்நாள் நிகண்டினை அவ்வப்போது புரட்டிப் பார்க்க வேண்டுவதில்லை. நிகண்டுகற்றார் அனைவரும் உலவும் நிகண்டராய் வாழ்ந்து வந்தனர்.”

27 The examples examined concern what Civan did to Irāvaṇaṇ, what Māl did to a wild lime tree, how a mountain was bent, etc.

Bibliography

1. Caṅmukam Piḷḷai, Mu. & Cuntaramūrṭti, I. (patippācīriyarkaḷ), 1990 & 1993, *Tivākaram*, 2 vol., Ceṅṅai Palkalaik Kaḷakam, Chennai.
2. *Cēntaṅ Tivākaram, Piṅkalam, Cūḷamaṅṅi*, 2004, Cānti Cātaṅṅā, Chennai.
3. Chevillard, Jean-Luc, 1996, *Le commentaire de Cēṅṅavaraiyar sur le Collatikāram du Tolkāppiyam*, Publication du Département d'Indologie N°84-1, Institut Français de Pondichéry & Ecole Française d'Extrême-Orient, Pondicherry.
4. Hüllen, Werner, 2004, *A History of Roget's Thesaurus, Origins, Development and design*, Oxford University Press.
5. James, Gregory, 1989, " A typological Scheme for the *Nikaṅṅu* Tradition ", *Journal of Tamil Studies*, N°36, Madras.
6. Knuth, Donald, E., 1993, *The Stanford GraphBase. A Platform for Combinatorial Computing*, ACM Press, Addison-Wesley Publishing Company, U.S.A.
7. *Piṅkalantai eṅṅum Piṅkala Nikaṅṅu*, 1968, Kaḷaka Veliyiṅṅu 1315, Tirunelvēli Teṅṅintiya Caivacittānta Nūṅpatippuk Kaḷakam, Chennai..
8. Subramoniam, V.I., 1962, *Index of Puranaanuuru*, Department of Tamil, University of Kerala.
9. Tāṅṅavarāyamutaliyār (patippācīriyar), 1840, *Cēntaṅ Tivākaram*, American Mission Press.
10. Subramanya Aiyar, V.M., Chevillard, J.-L., S.A.S. Sarma, 2007, *Digital Tēvāram. Kaṅṅiṅṅit Tēvāram* [CD-ROM], Collection Indologie n° 103, IFP / EFEO, Pondicherry.

சங்க இலக்கியத் தொடரடைவி

முனைவர் கா. உமாராஜ்

இணைப் பேராசிரியர், மொழியியல் துறை

மதுரைக் காமராசர் பல்கலைக் கழகம்

மதுரை - 625 021

k_umaraj@yahoo.com 94872 23316

இரா. அகிலன்

நிரலாளர், செம்மொழித் தமிழாய்வு மத்திய

நிறுவனம்

6, காமராசர் சாலை, சேப்பாக்கம், சென்னை - 5

akilan.rp@gmail.com 99657 34497

முன்னுரை

தொடரடைவு என்பது “ஒரு நூலில் அல்லது பதிப்பில்” பயன்படுத்தப்பட்டுள்ள சொற்களின் அகர வரிசைப்படுத்தப்பட்ட பட்டியலாகும். இப்பட்டியலில் சொற்களுடன் அவை இடம்பெறும் தொடர்களும் குறிப்பிடப்பட்டிருக்கும். தொடரடைவு என்பதை விக்கிபீடியா பின்வருமாறு பொருள்கொள்கிறது “A Concordance is an alphabetical list of the Principal words used in a book or body of work with their immediate contexts”. பைபிள், குர்ஆன், வேதங்கள், ஷேக்ஸ்பியரின் நூல்கள் போன்ற நூல்களுக்கு மட்டுமே தொடரடைவி உள்ளது. ஆனால் இதுவரை சங்க இலக்கியங்களுக்கெனத் தனியாக ஓர் தொடரடைவி உருவாக்கப்படவில்லை. இக்கட்டுரை சங்க இலக்கியங்களுக்கெனத் தொடரடைவி உருவாக்கவது பற்றி விவாதிக்கிறது.

தொடரடைவி

தொடரடைவி என்பது கொடுக்கப்பட்ட ஏதேனும் ஒரு சொல்லின் தொடரடைவை உருவாக்கித் தரக்கூடிய கணினி நிரல். தொடரடைவு என்பதைக் கணினியில் உள்ளீடு செய்யப்பட்டுள்ள ஒரு நூலிலிருந்து தரவுத் தளங்களை உருவாக்கும் முறை என்று வரையறுக்கலாம். இப்பணியைச் செய்துதரும் மென்பொருள் தொடரடைவி எனப்படுகின்றது. இம்முறையில் உருவாக்கப்படும் மென்பொருளில் விருப்பத்திற்கேற்பச் சொல் நிகழ் எண்ணிக்கை போன்ற பிற வசதிகள் செய்யப்பட்டுள்ளன. இவ்வாறு பெறப்படும் தொடரடைவிகளின் வெளியீடுகள் மற்ற மொழியியல் கருவிகளான சொல் மற்றும் இலக்கணத் திருத்தி, தேடு பொறிகள், மொழிசார்க் கருவிகள் போன்றவற்றுக்கு மிகவும் பயன்படக்கூடியன.

தொடரடைவிகளிலிருந்து பெறப்படும் தரவுகள் குறிப்பிட்ட சொற்கள் இடம்பெறும் தொடர்களைக் கொண்டிருக்கும். இந்தப் பட்டியல் உருவாக்கப்படும் தரவுத்தளம் தொடரடைவிகள் உருவாக்கியதாகவோ அல்லது வேறு வழியில் பெறப்பட்டதாகவோ இருக்கலாம். இவ்வாறு கணினியால் கையாளப்படும் வகையில் சொற்களின் தரவுத்தளங்களைக் கொண்டிருப்பவைதான் தரவகங்கள்(Corpus) எனப்படுகின்றன. இவ்வகைத் தரவகங்கள் சிறப்பாக உள்ள ஒரு மொழியில் சிறந்த கணினிசார் மொழிக் கருவிகளை உருவாக்கமுடியும். இந்தத் தரவகங்கள் ஒரு மொழியில் ஏற்படும் மாற்றங்களுக்கு ஏற்ப அடிக்கடி மேம்படுத்தப்படவும் வேண்டும். இவ்வகையான தரவகங்கள் கணினிசார் மொழிப் பயன்பாட்டுக்கு மிகவும் அடிப்படையாகும்.

தொடரடைவியின் முக்கியத்துவம்

இதுவரை கணினியில் சங்க இலக்கியங்களின் பயன்பாடு ஆரம்ப நிலையிலேயே உள்ளது. கணினி மற்றும் இணைய மொழி பயன்பாட்டில் பல நிலைகள் இனிமேல்தான் அடையப்பட வேண்டும். இவற்றைத் தொடரடைவிகள் மூலம் எளிதாக அடைய முடியும். இவை தொடரடைவு கணினிசார் பயன்பாடுகளுக்கு ஒரு

மொழியில் மிகவும் வலுவான அடித்தளமாக அமைய வாய்ப்பு உடையது. தொடரடைவிகள் அடிப்படை சொல் திருத்திகளிருந்து எந்திர மொழி மாற்றங்கள் வரை எதிர்கொள்ளப்படும் சிக்கல்களுக்குத் தீர்வளிக்கக் கூடியவை. ஏனெனில், கணினிமொழிக் கருவிகளில் ஒவ்வொன்றுமே சிறப்பான தரவுத்தள மேலாண்மையை அடிப்படையாக கொண்டது. இந்தத் தரவுத்தள மேலாண்மை சாதாரணத் தரவுத்தள மேலாண்மை மென்பொருள்களை மட்டுமின்றித் தொடரடைவு மற்றும் தரவகங்களை அதிகம் சார்ந்திருக்கிறது. ஏனெனில் இவ்வகைக் கருவிகளே பிராந்திய மொழிகளில் சிறப்பான முறையில் தரவுத்தள மேலாண்மையைத் தர முடியும். எனவே தொடரடைவு மற்றும் தரவகங்கள் சார்ந்த பணிகளில் நாம் காட்டும் ஈடுபாடு சிறப்பான கணினிமொழிக்கருவிகளின் உருவாக்கத்துக்குத் துணை செய்யும். மேலும் இதனால் தமிழ் மொழி கணினியில் விரைவில் தன்னிறைவு அடையும்.

தொடரடைவியின் பயன்கள்

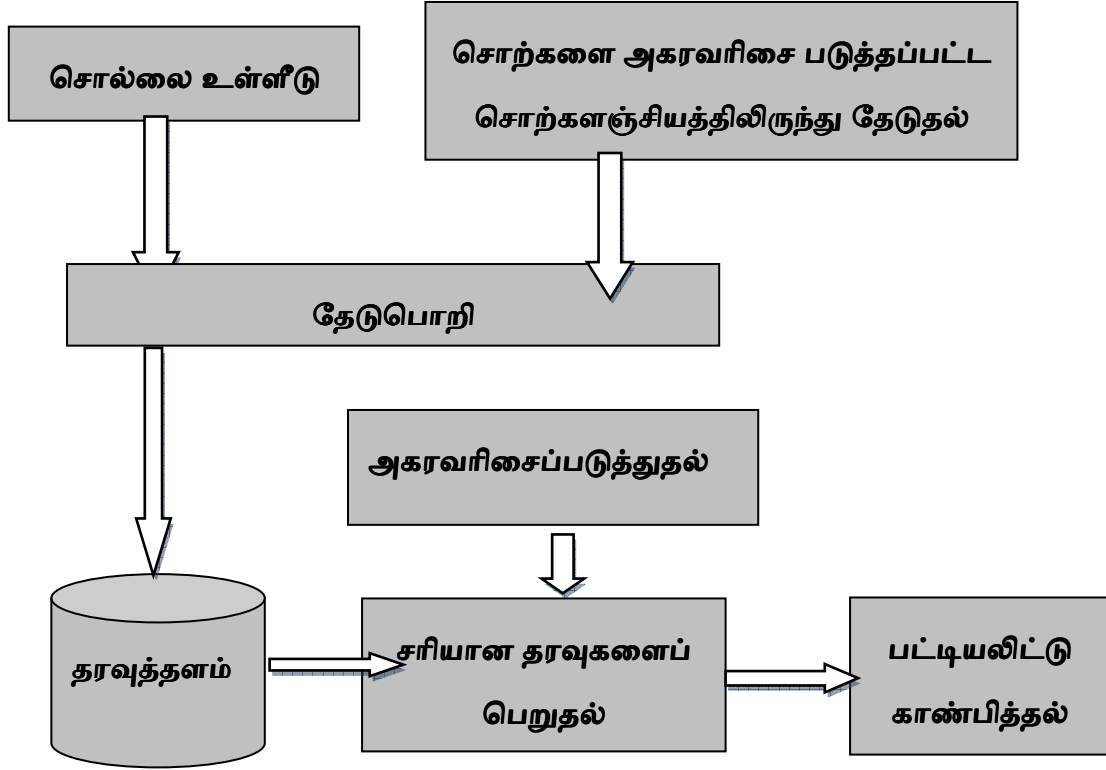
- ஒரு சொல் சங்க இலக்கியங்களில் எத்தனை முறை வருகின்றது என்பதைத் தொடரடைவி மூலம் அறியலாம்.
- ஒரு சொல் அதன் தொடரியல் சூழலுக்கேற்பப் பொருள் மாற்றம் பெறுவதைத் தொடரடைவி மூலம் அறியலாம்
- மொழியியல் விதிகள், கோட்பாடுகள் சரியானவைதானா என்பதைத் தொடரடைவி மூலம் அறியலாம்.
- வரலாற்று முறை இலக்கணம், வரலாற்று முறை அகராதி தயாரிப்பதற்குத் தொடரடைவி பயன்படுகிறது.
- சீர், உருபன், சொல், தொடர் முதலியவற்றைப் பிரிக்கப் பயன்படும் கருவிகளை உருவாக்குவதற்கும் தொடரடைவி பயன்படுகின்றது.
- வேர்ச் சொல்லாய்வுகள், பல்வேறு அகராதிகள், பொருட்களஞ்சியங்கள் தயாரிப்பதற்கும் தொடரடைவி பயன்படுகின்றது.
- கல்வி நுட்பவியல், மொழி நுட்பவியல் துறைகளுக்கும், பின்வரும் மென் பொருள் உருவாக்கத்திற்கும் தொடரடைவி பயன்படுகிறது.
 - ✓ மின் அகராதி
 - ✓ சொல் திருத்தி
 - ✓ இலக்கணத் திருத்தி
 - ✓ பேச்சு மொழியிலிருந்து எழுத்து மொழியை உருவாக்கும் கருவி.
 - ✓ எழுத்து மொழியிலிருந்து பேச்சு மொழியை உருவாக்கும் கருவி, மொழிபெயர்ப்பு கருவி.

தரவுகள்

பின்வரும் நூல்களிலுள்ள பாடல்கள் அப்பாடல் எண், தொடர் எண்ணுடன் தரவகத்தில் ஒருங்குறி முறையில் உள்ளீடு செய்யப்பட்டுள்ளன.

1.நற்றிணை	10.பொருநராற்றுப்படை
2.குறுந்தொகை	11.சிறுபாணாற்றுப்படை
3.ஐங்குறுநூறு	12.பெரும்பாணாற்றுப்படை
4.பதிற்றுப்பத்து	13.முல்லைப்பாட்டு
5.பரிபாடல்	14.மதுரைக்காஞ்சி
6.கலித்தொகை	15.நெடுநல்வாடை
7.அகநானூறு	16.குறிஞ்சிப்பாட்டு
8.புறநானூறு	17.பட்டினப்பாலை
9.திருமுருகாற்றுப்படை	18.மலைபடுகடாம்

மென்பொருள் அமைப்பு



சிக்கல்கள்

தொடரடைவி உருவாக்கும்போது ஒரு தொடரிலுள்ள சொற்கள் மட்டுமின்றி அத்தொடரின் பொருளும் மிகவும் இன்றியமையாதது. தொடரின் பொருளை முழுமையாக அறிந்தால் மட்டுமே சொற்களைத் தெளிவான முறையில் பிரிக்க முடியும். சங்க இலக்கியத்தில் எது சொல்; எது சொல் ஆகாது என்பதை வரையறுத்துக் கூறுவது சிக்கலானது. சில அறிஞர்கள் தனிச்சொல்லைக் கூட்டுச் சொல்லாகவும், கூட்டுச் சொல்லைத் தனிச்சொல்லாகவும் பிரித்துள்ளனர்.

உதாரணமாக "உரை திகழ் கட்டளைக் கடுப்ப மாச்சினை (குறு.192:4)" எனும் தொடரில் 'மாச்சினை' எனும் சொல்லில் "மா" எனவும் "சினை" எனவும் பிரிக்கலாம். இது மா மரத்தின் கிளை என பொருள் தரும். ஆனால் 'மாங்கனி நறுங்கடி கூட்டுவோம்' எனும் தொடரில் 'மாங்கனி' என்ற சொல்லை பிரிக்கக்கூடாது எனினில் இது கூட்டுச்சொல்.

V.I.Subramaniam also accepts this view and quote as follows "In Some places segmentation of the head words was problematic". The principles of segmentation adopted are: If the units A and B mean X and Y, A is listed with meaning x and B with y: If A and B mean only Z then A and B are listed as one word. In some cases A and B having the gloss Z may yield on etymological or metaphorical guess, an isolable gloss for A and B. For example in the form 'Vilaimakal' = 'prostitute', which is treated as a single word, it is possible to assign the meaning 'price' and 'women' to 'vilai' and 'makal', the prostitute is a woman who

sells her body. Such investigations are eliminated in plotting the meaning. The conclusion is if the word is lexicalized then it should not be splitted.

சங்க இலக்கியத் தொடரடைவியைப் பயன்படுத்தும் முறை

மேற்கண்ட தரவுகளைப் பயன்படுத்தி VB.Net மென்பொருளின் உதவி கொண்டு சங்க இலக்கியத்திற்கான தொடரடைவுக்கருவி உருவாக்கப்பட்டுள்ளது. இதில் இரண்டு வகையான உள்ளீடு கொடுக்கும் முறை பயன்படுத்தப்பட்டுள்ளது.

1. Text Box

2. Combo Box

- Text Box - ல் பயனாளர் சங்க இலக்கியச் சொல்லைத் தட்டச்சு செய்து பின்னர்த் தெரிவு செய்து அந்தக் குறிப்பிட்ட சொல்லுக்குரிய தொடரடைவைப் பெறலாம்
- Combo Box - ஐ பயனாளர் குறிப்பிட்ட சங்க இலக்கியச் சொல்லைத் தெரிவு செய்து
- அந்த குறிப்பிட்ட சொல்லுக்குரிய தொடரடைவைப் பெறலாம்
- குறிப்பிட்ட சொல் சங்க இலக்கியத்தில் எத்தனை முறை பயின்று வந்துள்ளது என்பதையும் அறியலாம்.
- அந்தச் சொல் இடம்பெற்றுள்ள நூலின் பெயர்,பாடல் எண், அடி எண் ஆகியவற்றையும் அறியலாம்.
- மேலும் அந்தச் சொல் இடம் பெற்றுள்ள பாடல் அடியின் முழு பாடலைப் பெறவேண்டும் எனில்

Row Header ஐ தெரிவு செய்தால் அதற்கான முழுப் பாடலைப் பெற இயலும்.

முடிவுரை

சங்க இலக்கியங்களுக்கெனத் தொடரடைவி உருவாக்கும்போது எவ்வகையான சொற்பிரிப்பு முறையை பின்பற்றுவது என்பதை முடிவு செய்தல் வேண்டும். தற்போது உருவாக்கியுள்ள தொடரடைவியில், மொழியியல் அடிப்படையிலான சொற்பிரிப்பு முறையில் குறிப்பாக, தாமஸ் மால்டன், மாதையன் ஆகியோர் உருவாக்கிய சொற்பிரிப்பு நெறிமுறையையே பயன்படுத்தி ஆசிரியர் தொடரடைவியை உருவாக்கியுள்ளார்.

Nature of Tamil Data and Computing

தமிழ்த் தரவின் தன்மையும் கணிப்படுத்தும்

Dr. P. David Prabhakar
Associate professor of Tamil,
Madras Christian College, Chennai-600 059
tamilprofessor@gmail.com

Abstract

Both computer scientists and linguists are working towards developing natural language tools, resources and real- world NLP applications. Both aspire to comprehend the aspects of the language that makes analyzing computationally tractable. Understanding the nature of language is as important as understanding the computing environment. Mutual understanding in this regard will lead us towards expected outcome.

This paper describes the nature of Tamil data in detail with a view to the requirements of computing in terms of...

1. Spoken vs. written nature of Tamil data
2. Sequential nature of Tamil data
3. Hierarchical nature of Tamil data
4. Multi dimensional nature of Tamil data
5. Integrated nature of Tamil data
6. Multi lingual nature of Tamil data

மொழித் தரவின் பொதுப் பண்புகள் மூளையின் உயிரியல் அடிப்படைகளைச் சார்ந்துள்ளன. கணினியின் செயற்பாடு மின்னணுவியல் நுட்பத்தைச் சார்ந்துள்ளது. மொழித் தொழில் நுட்பமானது பகுப்பாய்வு ஒருங்கு, வெளியீட்டு ஒருங்கு, ஊடாட்ட ஒருங்கு ஆகியவற்றைக் கொண்டது. மொழித் தரவுகளைக் கணினி திறம்படக் கையாள்வதற்கும் தேவையான வெளியீட்டைப் பெறுவதற்கும் கணினி அடிப்படைகளோடு தமிழ்த் தரவின் பன்முகத் தன்மைகளை ஆய்ந்து அறிவது இன்றியமையாதது.

இதுவரை நிகழ்ந்துள்ள தமிழ்க் கணிப்படுத்த முயற்சிகளையும் போக்குகளையும் பற்றிய மதிப்பீடு, வருங்காலத் தமிழ்க் கணிப்படுத்தம் பெரிதும் மொழியியல் அறிவை உட்படுத்த வேண்டிய தேவையை வற்புறுத்துகிறது. இதனால், மொழியின் பன்முகப் பண்புகளைத் தனித்தனியே விவரிப்பதன் வழி, அவற்றைக் கையாள்வதற்குரிய கணினியச் சாத்தியப்பாடுகள் குறித்த விவாதத்திற்கு இக்கட்டுரை வழிவகுக்கிறது. தமிழ்த் மொழித் தரவுகள் உலகளாவிய பொது மொழியின் தன்மைகளைக் கொண்டிருப்பதோடு, தமக்கென தனித்த சில தனித்தன்மைகளைக் கொண்டிருப்பதும் இங்கு கவனத்திற்கொள்ளப்பட்டுள்ளது.

செம்மொழியாகவும் நவீன மொழியாகவும் விளங்குவது தமிழ். நெடிய வரலாற்றினைக் கொண்டிருப்பதால் சொல்வளம் மிகுந்த மொழி. வழக்கிழந்த கூறுகளும் புத்தாக்கங்களும் இம் மொழியில் உண்டு. வட்டாரம், சாதி, தொழில் சார் வேறுபாடுகளும் தமிழில் உண்டு. பேச்சு, எழுத்து எனும் இரட்டை வழக்கு நிலையைக் (Diglossia) கொண்டிருப்பது தமிழ். பிற மொழிச் சொற்களின் பயன்பாடும் தமிழ்ப் பணுவல்களில் தவிர்க்க இயலாது இடம் பெறுகிறது.

மொழித் தரவின் தன்மைகள்

பேச்சுத் தரவும் எழுத்துத் தரவும்

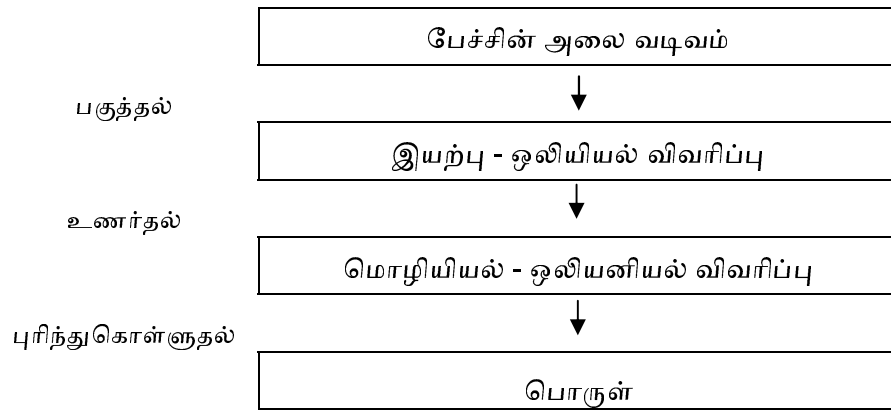
கணினியில் உள்ளீடு செய்யப்படும் தமிழ்த் தரவு பேச்சாகவோ எழுத்தாகவோ அமைகிறது. விசைப்பலகை வாயிலாக உள்ளீடு செய்யப்படும் தரவைக் கணினி எளிதாகவும் முழுமையாகவும் ஏற்கிறது. ஆயினும் கையெழுத்துப் பணுவலைப் புரிந்து கொள்வதிலும் ஒளி வருடியைப் பயன்படுத்தி பெறும் மொழி ஆவணத்தைப் புரிந்து கொள்வதிலும் கணினிக்குச் சிக்கல்கள் எழுகின்றன. ஒளியியலை அடிப்படையாகக் கொண்டு தமிழ் எழுத்துக்களை அடையாளம் காணும் முயற்சிகள் நடைபெற்றுள்ளன. தமிழ் எழுத்துகளின் படிநிலைக் குழு விவரிப்பில் அமையும் துவக்கக்கோடு மட்டம், பகுதிக்கூறு மட்டம் மற்றும் முழு எழுத்து மட்டம் ஆகியவற்றின் வரைபட அமைப்புகளில் சில இறுக்கிகளைப் (Constrains) பயன்படுத்தி எழுத்துக்களை இனம் காணும் ஆய்வுகள் நிகழ்த்தப்பட்டுள்ளன.

மேலும், யுனிகோடுக்கு முந்தைய கணினியச் சூழல் பன்மொழித் தரவுகளைக் கையாள்வதிலும், குறியீட்டு முறைகளிலும் சிக்கல்களைச் சந்திக்க வேண்டியிருந்தது. தமிழ் தொடர்பாக இன்றைய சிக்கல்கள் தீர்ந்து வருகின்றன. பேச்சுத் தரவை இரு வகைகளில் கணினி கையாள வேண்டியுள்ளது.

அ. பேச்சொலிகளைக் கண்டறிதல் (Speech Recognition)

ஆ. பேச்சொலிகளை இணைத்தல் (Speech Synthesis). எழுத்துத் தரவை ஒப்பிடும் நிலையில் பேச்சுத் தரவைப் புரிந்துகொள்வதில் கணினியின் திறனை மிகுதியும் பயன்படுத்த வேண்டியுள்ளது.

பேச்சுத் தரவுகள் அலை வடிவில் கணினியால் பெறப்படுகின்றன. இவை இயற்பு - ஒலியியல் விவரிப்புக்கும் (Acoustic - Phonetic) மொழியியல் - ஒலியனியல் விவரிப்புக்கும் (Linguistic - Phonemic) உட்படுகிற நிலையில் தமிழ் ஒலிகளை அடையாளம் காண இயலும். இதனைப் பின்வருமாறு விளக்கலாம்.



மேலும் பேச்சுப் பணுவலில், மீட்பகுப்புக் கூறுகளாக அமையும் அழுத்தம், சுரம், விட்டிசை, நெடுமை முதலியவற்றைக் கணினி கையாள வேண்டியுள்ளது. சூழல், ஆள் சார்ந்து ஒரே சொல் ஒன்றுக்கும் மேற்பட்ட முறையில் உச்சரிக்கப்படுதல், உச்சரிப்பில் நிகழும் ஏற்ற இறக்கம், உரத்து ஒலித்தல், கால நீட்சி போன்ற கூறுகள், பேசுவரின் விருப்பு, வெறுப்பு, மகிழ்ச்சி, இகழ்ச்சி போன்ற மனப்பாங்கு உணர்த்தும் கூறுகள் ஆகியவற்றைக் கணினி கையாள்வது கடினம். தமிழில் காணப்படும் குரலிசையை (Intonation) ஆய்வு செய்த ரவிசங்கர் (1987) 67 வகைகளைக் கண்டறிந்துள்ளார். மேலும், பேச்சில் இடம்பெறும் உயிர்ப்பு ஒலி (Aspirated) அதிர்வுடை ஒலி (Voiced) போன்ற நுட்பமான ஒலியியல் வேறுபாடுகளை அடையாளம் காண்பதும் கணினியின் ஏற்புத் திறனுக்கு சவாலாக அமைவதாகும்.

வரிசைமுறைப் பண்பு (Sequential nature)

மொழித் தரவுகள் பொதுவாக வரிசை முறையில் அமைகின்றன. ஒலிகளும் (அல்லது எழுத்து) சொற்களும், தொடர்களும் பொருள் தரும் முறையில் ஒன்றையொன்று அடுத்தார்ந்து வருகின்றன. மொழித் தரவின் வரிசை முறை மாறினால் வழி அல்லது பொருள் வேறுபாடு தோன்றக்கூடும். தமிழில் எழுத்துக்களின் வருகை, மயக்கம், ஒரு சொல் பிற மொழி அலகுகளை ஏற்கும் இட வரையறை போன்றவை தனித்தன்மை வாய்ந்ததாகும். எழுத்து வடிவில் அமையும் மொழித் தரவுகள் எழுதப்படும் பரப்பின் அளவு கருதி மடித்து எழுதப்படலாம்; பத்தி பத்தியாகவும் பக்கம் பக்கமாகவும் எழுதப்படலாம். பக்கங்களின் வலப்புற ஒழுங்கு கருதி, ஒரு சொல் நீர்மை கொண்ட மொழி அலகுகள் சிறு கோடிட்டு (Hyphen) பிரித்து எழுதப்படுவதும் உண்டு. செய்யுள் வடிவில் அமைந்த மொழித் தரவுகள் யாப்புக் கட்டுப்பாடுகளுக்கிணங்க, 'அடி' எனும் பகுப்பைக் கொள்ளும். மொழித் தரவின் இத்தகைய வரிசைமுறைப் பண்புகளைக் கணினியில் செயல்படும் தொகுப்பான், சொல்லாளர் போன்றவை எளிதாகக் கையாளுகின்றன. ஆயினும் மொழியின் பிற பண்புகளைக் கையாள்வதற்கு இவை போதுமானதன்று.

படிநிலைப் பண்பு (Hierarchical nature)

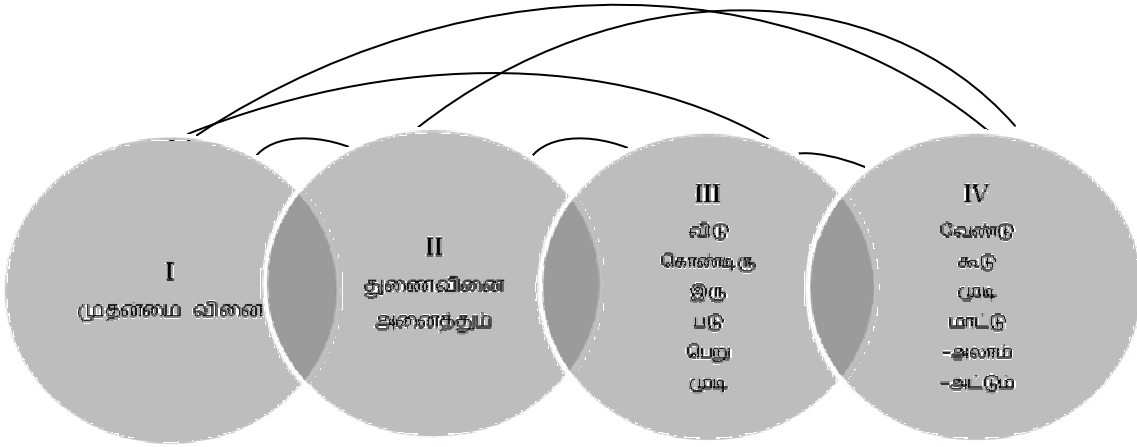
மொழி அலகுகள் வரிசை முறைப்பண்புகளைக் கொண்டிருப்பதோடு படிநிலைப் பண்புகளையும் கொண்டிருக்கின்றன. தொடரியல் அணுகுமுறை மொழி அலகுகளைப் படிநிலை கொண்டதாக அணுகுகிறது. தலைமை அலகு, சார் அலகு ஆகிய நிலைகளையும் புதிய இலக்கணக் கோட்பாடுகள் சுட்டுகின்றன. சிட்னி லாம்ப் (Sydney Lamb) மொழியின் அமைப்பு பல்வேறு அடுக்குகளை (Structure layer or Strata) கொண்டதாக மொழி இலக்கணத்தை அணுகுவது இங்குக் கருத்தத்தக்கது. மொழியின் படிநிலைப்பண்பைக் கையாள சிறப்பு மொழிகளும் (SGML) இருந்து வருகின்றன.

ஒரு பணுவலில் அமையும் ஒன்றுக்கு மேற்பட்ட உரைகளுக்கிடையே அமையும் பொருள் தொடர்பை விளக்க நேரலாம். இவற்றைக் கருத்தாடல் (Discourse) நிலையிலேயே அணுகமுடியும். இவை மொழி அமைப்பு எல்லைகளைக் கடந்தாகவும் உள்ளன.

அகராதிகளில் இடம்பெறும் மொழித் தரவுகளும் படிநிலை அமைப்பைக் கொண்டுள்ளன. தலைமைச் சொல், இலக்கண வகை, வேர்ச்சொல், பொருள்கள், எடுத்துக்காட்டுகள் ஆகியன ஒவ்வொரு அகராதி பதிவிலும் இடம்பெறக் கூடும். இத்தகைய தரவுகளை வரிசைமுறையில் அமைந்திருப்பதாகக் கொள்ள இயலாது. இதுபோன்றே இணைச் சொற்கள் எதிர்சொற்கள் ஆகியனவும் படிநிலைப் பண்புகளைக் கொண்டவை. வாக்கியங்களில் அமையும் அலகுகள் மட்டுமின்றி, ஒரு சொல்லமைப்பில் இடம் பெறும் பல்வேறு உருபுகளும் வரிசைமுறைப் பண்புகளைக் கொண்டிருப்பதோடு, அடுக்குமுறைப் பண்புகளையும் கொண்டுள்ளன. தமிழ்த் துணைவினைகள் வரிசைமுறையிலும் அடுக்குமுறையிலும் கொள்ளும் உறவு எடுத்துக்காட்டுக்காகக் கீழே கொடுக்கப்பட்டுள்ளது. (ப.டேவிட் பிரபாகர் 2002)

பன்முகப் பண்பு (Multi dimensional nature)

ஒவ்வொரு சொல்லும் இலக்கணப் பண்புகளைக் கொண்டிருப்பதோடு பொருளையும் உணர்த்துகின்றன. ஒவ்வொரு மொழி அலகையும் ஒலி, அசை, உருபு, தொடரியல் போன்ற பல்வேறு நோக்குகளில் கையாள வேண்டியுள்ளது. ஒரு சொற்றொடரில் அமையும் பெயர்கள் எழுவாயாகவோ, செயப்படுபொருளாகவோ செயல்படலாம். ஒரு சொல் நேர்ப் பொருளையோ, குறிப்புப் பொருளையோ உணர்த்தக்கூடும். மரபுத் தொடர்கள், மங்கலம், இடக்கரடக்கல் போன்றவற்றைக் குறிப்பிட்ட மொழி சார்ந்து கையாள வேண்டியுள்ளது. இத்தகைய பண்புகள் மொழி அலகுகளின் பன்முகத் தன்மையைக் காட்டுகின்றன.



பேச்சுத் தரவில் எழுத்து / சொல் சார்ந்து வெளிப்படும் சுரம், குரல், விட்டிசை, அழுத்தம், நெடுமை போன்ற மீப்பகுப்புக் கூறுகளும் மொழியின் பன்முகப் பண்பைச் சுட்டுவன. செய்யுளில் அமையும் எதுகை, மோனை முதலிய ஓசையியல்புகளும் மொழித் தரவின் பன்முகத் தன்மையைக் காட்டுகின்றன. மொழித் தரவின் பன்முகப் பண்புகளை வரையறுத்துத் தர இயலுமானால் அவற்றைக் கணினியில் கையாள இயலும்.

ஒருங்கிணைந்த தன்மை (Integrated nature)

மொழித் தரவின் பன்முகப்பண்புகள் ஒன்றையொன்று சார்ந்து இயங்குவன. ஒவ்வொரு மொழி அலகையும் எழுத்தியல், சொல்லியல், தொடரியல் உள்ளிட்ட பல்வேறு நோக்குகளுக்கு ஒரே சமயத்தில் உட்படுத்த வேண்டிய தேவை உள்ளது. பெயர்ச் சொற்கள் காட்டும் திணை, பால், எண், இடம் ஆகியன தமிழ்த் தொடரமைப்பில் இன்றியமையாத கூறாகச் செயல்படுகின்றன. பெயர்ச் சொற்கள் வினையைச் செய்பவராகவோ (agent) வினைக்கு உட்படுபவராகவோ (patient) வினையால் பயன்பெறுபவராகவோ (beneficiary) செயலை உணர்பவராகவோ (experiencer) செயலைச் செய்யும் கருவியாகவோ அமையலாம். இவற்றில் அடிப்படையிலேயே தொடரில் வேற்றுமை உறவுகள் அமைகின்றன. சில வினைகள் இயற்கையாகவே செயப்படுபொருள் குன்றிய வினைகளாக உள்ளன. இத்தகைய பண்புகளையும் கணினி

கையாள இயலும். ஆயினும், மொழியின் சமுதாய பண்பாட்டுச் சூழல் சார்ந்த பரிமாணங்களைக் கையாள்வது கணினியின் ஆளுகைக்குச் சவாலாக அமையும்.

பன்மொழித் தன்மை (Multi lingual nature)

தமிழ்ப் பனுவல்களில் பல்வேறு சூழல்களில் பிறமொழிச் சொற்களின் பயன்பாட்டைக் காணலாம். பிறமொழிச் சொற்கள் பிறமொழி எழுத்துருக்களால் ஆளப்படும் பொழுது அவற்றைக் கையாள்வதில் கணினிக்குச் சிக்கல் ஏற்படக் கூடும். அண்மைக் காலத்தில் பரவலாக அறிமுகமாகியுள்ள யுனிகோட் முறை தமிழ் உள்ளிட்ட பல்வேறு மொழிகளின் குறியீட்டிற்கு நிலையான இடம் வழங்கியிருப்பதால் பன்மொழித் தரவையையும் இனிக் கணினியால் கையாள இயலும். தமிழ்ச் சூழலில் காணப்படும் ஒன்றுக்கும் மேற்பட்ட குறியீட்டு முறைகள் (encoding) விசைப்பலகை அமைவுகள் (key board layout) ஆகியவற்றில் சீர்மை ஏற்பட்டு வருகிறது.

இதுகாறும் வரையறுக்கப்பட்ட நோக்கங்களுக்காகவே தமிழ்த் தரவுகள் கணினி வழி கையாளப்பட்டுள்ளன. இணையத்தில் ஏற்பட்டுள்ள தமிழ்ப் பெருக்கம், புதிய இலக்குகள் ஆகியவற்றுக்குத்தக, மொழித் தொழில்நுட்பம் சார்ந்து தமிழ்த் தரவின் பன்முகப் பண்புகளை விவாதிப்பதும் வரையறைப்படுத்துவதும் வருங்கால தமிழ்க் கணிப்படுத்த முயற்சிகளுக்குத் துணைபுரியும்.

துணை நின்றவை

Alan Garnhan	1998	Artificial Intelligence : An Introduction, Rutledge, London,
Christopher, S Butler (Ed)	1992	Computer and Written Texts, Black Well, Oxford, UK
David Prabhakar, P.	2004	'Tamil Hyphenator' , International Tamil Internet Conference, Singapore.
	2009	'Computer Analysis of Tamil Verb forms, International Tamil Internet Conference, Germany
John Lowler and	1998	Using Computers in Linguistics: Practical Guide, Rutledge,
Helen Amistary Dry (Eds)		London 1998
Rushan Mitkov	2003	Oxford Hand book of Computational Linguistics, Oxford, UK

FaceWaves : A Tamil Text to Video Framework

Madhan Karky, T V Geetha & Ravi Varman

{madhankarky@gmail.com, tv_g@hotmail.com, rvmsdh@yahoo.co.in}

Department of Computer Science & Engineering

College of Engineering Guindy

Anna University

Abstract

This paper presents FaceWaves, a framework for interactive information interchange for Tamil Internet and mobile users. A text-to-video subsystem for generating faces and animating them purely based on textual descriptions is proposed. Using textual descriptions to create a face by describing the visual features of a person and animating their face to speak out given information, can provide an efficient means of storing/transferring an animated video. The Text-to video subsystem comprises of a morphological analyser, ontology of facial features and expressions, Tamil text to speech, lip synchroniser, and emotion handler. To best of our knowledge, this is the first framework proposed for interactive information interchange with sophisticated language technology tools such as text to speech, text to video and voice to text. This framework facilitates storing and transferring video files over a network as plain text and converting the text to a video with a lightweight local client. This paper describes various components of the framework and shows the results from text-to-face generation module. The paper concludes discussing the results, opening a new area for the Tamil computing research community.

Introduction

Increasing number of Internet users, exponentially growing content and limited available bandwidth [1] has always been a problem to the Internet community. The number of information sources in Tamil and the number of Tamil users who contribute and consume is increasing every hour with the advent of blogs, microblogs and social networks. Almost every Tamil newspapers around the world now have their own portals feeding news and articles in Tamil every hour. We see bots collecting news from multiple sources and summarising the news. Auto-Journalists are already collecting sports scores from a website and generating a full fledged human like report about the match discussing performance of players purely based on numbers. Tamil is very soon going to adapt to these new growing technologies.

We foresee the need for human like auto-reporters who can read out a given news article, or can summarize any information source. We foresee the need for virtual faces that can listen to user queries and respond with answers. We foresee a system that can create a set of characters, backdrops and animate them from plain Tamil textual descriptions. In this paper we propose FaceWaves, a framework that enables creating such videos that can be transferred over any network as plain text and converted to video using a lightweight local client. This framework provides an ideal platform for transferring videos across world wide web and also in mobile networks.

This paper is organised into five sections. The second section discusses the background and motivation of this paper with related literature. The FaceWaves framework and components are given in section three. Section four provides the results from our face generator module of the FaceWaves framework in detail with snapshots from our GUI. The fifth section concludes the paper with our work in progress and future research in this domain.

Background

Tat Seng et al., proposed a animation sequencing method from text which tries to build 3D animations from a manual [2]. The proposed system tries to capture the concepts from the manual and extracts the features to provide the user with the option of choosing between textual or graphical modes. Narichika et al., in 2009 proposes a similar system that generates a movie from a script to a TV like program [3]. In this work the authors describe a framework that creates agents and a prototype that demonstrates controlling the generated agents. We find this work closely related to our proposed FaceWaves framework with respect to creating agents from text. A few Tamil Text-to-speech systems exist [2,6,7] and FaceWaves uses an improved variation of a text to speech system presented in [2]. Tamil speech to text has various proposals such as [3] and a fully working system yet to be demonstrated. Detecting emotions from a given text has been carried out in different languages [9-11] and the FaceWaves framework proposes an emotion identifier and applying emotions on a computer-generated face.

FaceWaves Framework

Face Waves architecture comprises of three main components. Information System, Wave Processor and Language Tools. An Interface Manager will be developed as a web tool, integrating all services to be accessible by clients. The components of the system are described briefly in the following sections.

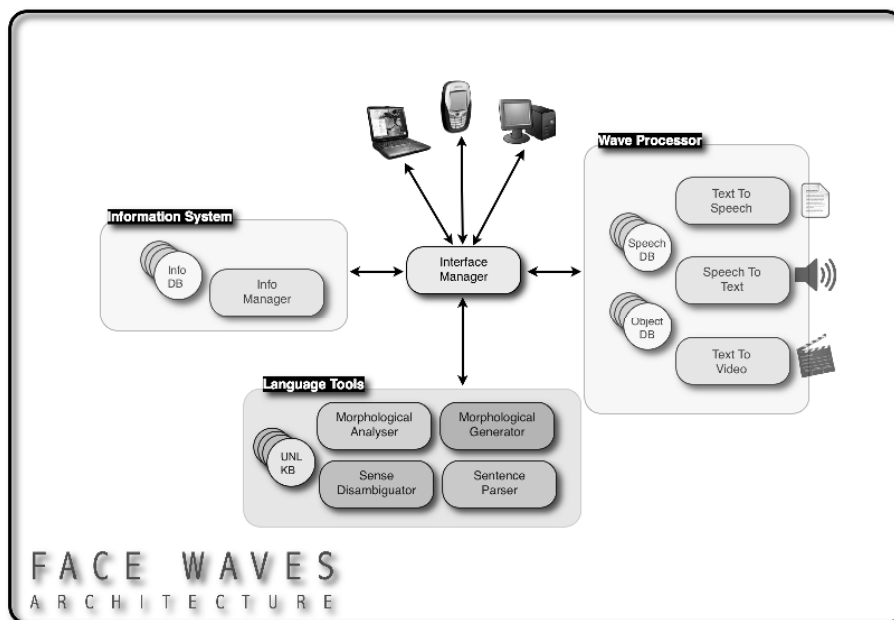


Fig 1 : FaceWaves Framework

a. Information System

Information System is responsible for crawling textual information on various topics over a structured information architecture such as Wikipedia(Tamil). The crawled information is processed and conceptually indexed using CoRe, a concept and relation based indexing system for large text collections. One key advantage of such a method would be to have a language independent indexing. The Information Manager will be responsible for crawling, processing, enconverting Tamil text to Universal Networking Language graphs, indexing the concept graphs, ranking documents based on concepts, retrieving and formatting results as requested by the Interface Manager.

b. Language Tools

Language Tools subsystem offers Tamil language dependent services such as morphological analysis, generation, named entity recognition, word sense disambiguation and sentence parser. These tools will be integrated into a system as a service for Wave Processor and Information System.

c. Wave Processor

Wave processor comprises three major units. Text to Speech, Speech to Text and Text to Video. These units work independent of each other and output of one unit can be piped as input to the other. The key responsibility of this system is to provide interaction elements (speech, agents, and agent animations) for the information interchange service provided by system. The wave processor makes use of language tools such as Tamil Analyser and Generator for various applications. The Wave Processor will be designed as a separate sub-system that can be plugged into different standalone, web or mobile applications.

d. Interface Manager

The interface manager acts as the central control object for FaceWaves. It holds the responsibilities of receiving queries from clients to create agents, displaying agents, collecting information as voice queries and sending to Wave Processor and converting the wave to text and sending the text to Information system to retrieve information and resend the information to Wave Processor to convert them to speech and agent animation and finally to play the animation video on the client's interface.

Text To Video Subsystem

The Text to Video subsystem of FaceWaves framework, as depicted in figure 2, comprises of a Document Processor, Face Generator, Backdrop Selector, Emotion Processor and a Movie Manager.

A plain text, semi-structured script file comprising of character descriptions, scene descriptions and dialogues is processed by Document processor to extract the corresponding sections and formats them structurally and routes the corresponding formatted information to Face generator, Backdrop selector and Emotion Processor. The face generator uses a face description ontology to describe dimensions of different parts of the human face for the given description. Backdrop selector uses a background library and analyses the scene description to rank and choose appropriate background for the given scene. Emotion Processor analyses the text and tags the dialogues with appropriate emotions using an Emotion ontology. The Face description, selected backdrop and annotated dialogues are then sent to the Movie manager which produces a plain text movie description file which can be sent back to the client. A lightweight movie player in the client on receiving the Movie Description, processes the description and

uses a Tamil Text to Speech module and synchronises the lip movements with the generated speech and face with appropriate emotion with the selected backdrop. The Movie Player generates a full length running animation movie with synchronised subtitles.

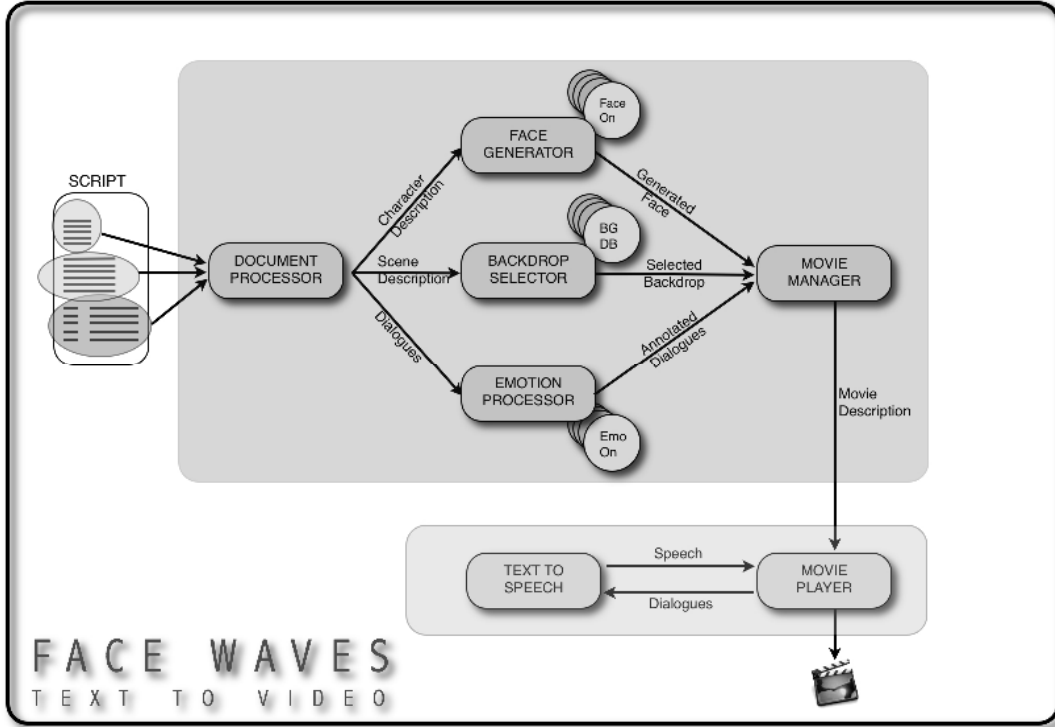


Fig 2 : Text to Video Subsystem

Results : Face Generator

This section discusses the Face Generator module and how plain text Tamil descriptions are converted to a human face. The Face Generator module as explained in the previous section uses a face description ontology along with a Morphological Generator.

A sample face description can be in a free flowing form such as

ஆதி அடர்த்தியான புருவங்களும் நீளமான மூக்கும் அகண்ட கண்களும் சிறிய உதடுகளும் கொண்டிருந்தான்

aadhi adarthyana puruvangaLum neeLamaana mookkum akaNda

kankaLum siRiya uthadugaLum koNdirunthaan

Aadhi has thick eyebrows, long nose, big wide eyes and small lips

The image generated for the face description provided above is provided in figure 3 as a snapshot from our text-to-video GUI.

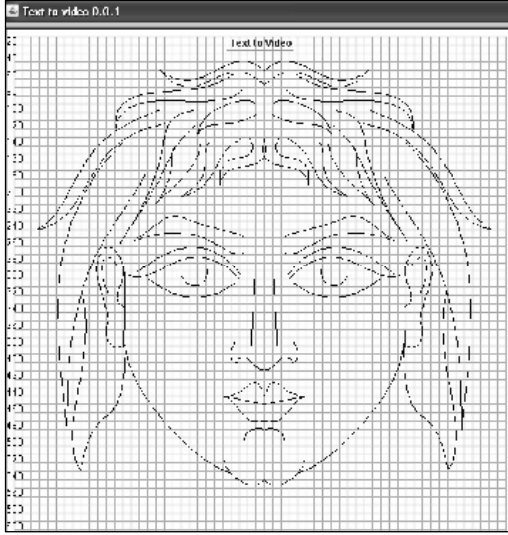


Fig 3 : Face Generator Snapshot 1

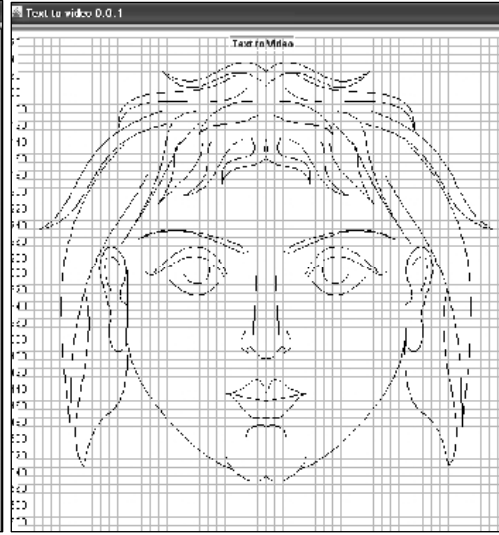


Fig 4 : Face Generator Snapshot 2

The snapshot in Figure 4 was generated for the following face description.

தியா சிறிய மூக்கும் மெலிதான புருவங்களும் பெரிய கண்களும் சிறிய உதும் கொண்டவள்.

thiyaa siRiya mookkum melidhaana puruvangaLum periya kaNkaLum siRiya uthadum koNdavaL

Dhiyaa has a small nose, thin eyebrows, big eyes and small lips.

The face generator module was developed in Java. The generator gets the face description text as input and uses a morphological analyser, sentence parser and face description ontology to retrieve the facial features from the given descriptions. A neutral face is initially described as a collection of objects one for each part of the face. The member variables of each object, which define the dimension of the corresponding part of face, are modified based on the adjectives describing each part in the face description.

Conclusion and Future Work

The current face generator module does not take into account the descriptions for hair, skin colour and texture, moustache, special marks or any three dimensional descriptions. Including these features in the face generation module will be our future work along with integrating the face generation module with adding expressions to the face based on the emotion identified from the dialogue. This paper provides the overall FaceWaves framework and describes one particular subsystem, Tamil text-to-video, and one key module of the subsystem, Face Generator. We believe that this paper will open numerous research problems for the Tamil computing community.

References

1. Anandan, R. Parthasarathi, and Geetha, *Morphological Analyser for Tamil*. ICON 2002, 2002.
2. Karky, M., et al. *Tamil Voice Engine*. in INFITT. 2001. Malaysia.
3. Lakshmi and H. Murthy. *A Syllable based continuous speech recognizer for Tamil*. in ICSLP. 2006. Pittsburgh: Interspeech.
4. Marketing, M., *World Internet Statistics*, <http://www.internetworldstats.com/stats.htm>. 2009.
5. Narichika, H., et al., *User-Definable Rule Description Framework for Autonomous Actor Agents*, in *Proceedings of the 13th International Conference on Human-Computer Interaction. Part III: Ubiquitous and Intelligent Interaction* %@ 978-3-642-02579-2. 2009, Springer-Verlag: San Diego, CA. p. 257-266.
6. Rama, J., et al., *A Complete Text To Speech System in Tamil*. IEEE, 2002.
7. Rao, N., et al. *Text-to-speech synthesis using syllable-like units*. in *National Conference on Communications*. 2005. Kharagpur, India.
8. Tat-Seng, C. and L. Thiam-Beng, *From Text Description to Animation Sequences*, in *Proceedings of the Computer Animation* %@ 0-8186-7588-8. 1996, IEEE Computer Society. p. 175.

Context Based Information Search for Thirukural

N.Ilakiaselvan

M.E - Software Engineering

CEG Anna University, Chennai-25

ilakiaselven@gmail.com, ilakiya_2909@yahoo.co.in

Abstract

ThiruKural is a discourse on the art of living, a set of healthy principles of guidance for the variety of segments of the civilization for a pleasant-sounding combined living. Each episodes and couplet(kurals) in Thirukural are related to the real time world. In the Tamil language "Thiru" means "holy" or "sacred," and "Kural" means anything that is brief or short. This paper focuses on context based searching for Thirukural and describes the techniques of Natural Language Processing (NLP) which is to design and build software that will analyze and apply in both understanding and generating natural languages that humans use naturally. A fundamental phenomenon of natural language is the variability of semantic expression. The system should understand the short story of some sentences like paragraph and result kurals with given relative meaning. Based on the given context, it retrieves the information of couplets (kurals) with the ranking priority. The system identifies the relative terms and using the term frequency calculates the weightage for each terms.

Keywords: Thirukural, Natural Language Processing, Couplets, Semantic expression, Term frequency.

Introduction

In this information search is based on the Natural Language Processing. NLP is a field of computer science and linguistic concerned with the interactions between computers and human natural languages. Two fields of NLP are: Natural language generation, system converts information from computer database into readable language and Natural language understanding, system converts human language into computers known format. The system participating in this competition must do something more than the system from other NLP competitions: to prove capabilities of understanding how language works.

Same meaning can be expressed by, or inferred from different texts mapping between language expressions and meanings. Humans use different expressions to convey the same meaning. Therefore, numerous NLP applications, such as, Question Answering, Information Extraction. Summarization require computational models of language that recognize the semantic approach. Trying to capture the major semantic inferences needed to understand equivalent semantic expressions. There are several levels of the meaning of the texts, ranging from shallow level to deep one. But, it is still difficult to make a consensus on how to describe the deep meaning.

In Question and Answering, an indexing is considered as the predictive annotation. Text Meaning Representation forms the Tamil dictionary which contains extended relative meanings. Information Retrieval must be started when the queries enters for searching couplets.

Literature survey

In Natural Language Processing, denotes as "Understanding" language means, among other things, knowing what concepts a word or phrase stands for and knowing how to link those concepts together in a meaningful way [6]. Basically Communication between Human and the system starts from Question and answering [3] and the main goal of QA is, whenever the user types a question, system must produce the correct answer. The Challenge is to analyze the questions, gather information and present the answer.

Question and answering is also related to text summarization technology [4], it contains short questions and answers. Sometimes there will be no answer for questions or multiple answers for questions, which have no answer object in documents to a given question or there are many answer objects. When a new question comes up, the system compares it with all the questions in the library and finds the most detailed standard question to match it. Then the corresponding answer is returned to the user [2].

While in Information Retrieval, An efficient indexing mechanism is normally used to quickly retrieve the information. And [7] Question Answering and information Retrieval are the two candidates of Natural Language Processing. An Indexing is used for searching, or finding the respective information. And the meaning represents the TMR language [1], which is the knowledge representation system for representing text meaning. An Onomasticon, or lexicon is nothing but the collection of proper names and terms; a dictionary.

Thirukural plays an important role of our human life[8] and the Greatness of Thirukural is a precious gem among the classics, unique in the deliverance of code of conduct to the mankind to follow for all time to come. This poem consisting of 133 sections of 10 couplets each which was predictable as a masterpiece of ancient literature in Tamil in its own times, has stood the test of history and is established by posterity as a decisive work which has predisposed the thoughts of man throughout the centuries. It is not only of great artistic and stylistic literary value, but also a direct to the art of living with pieces of precious wisdom.

How the search engine works

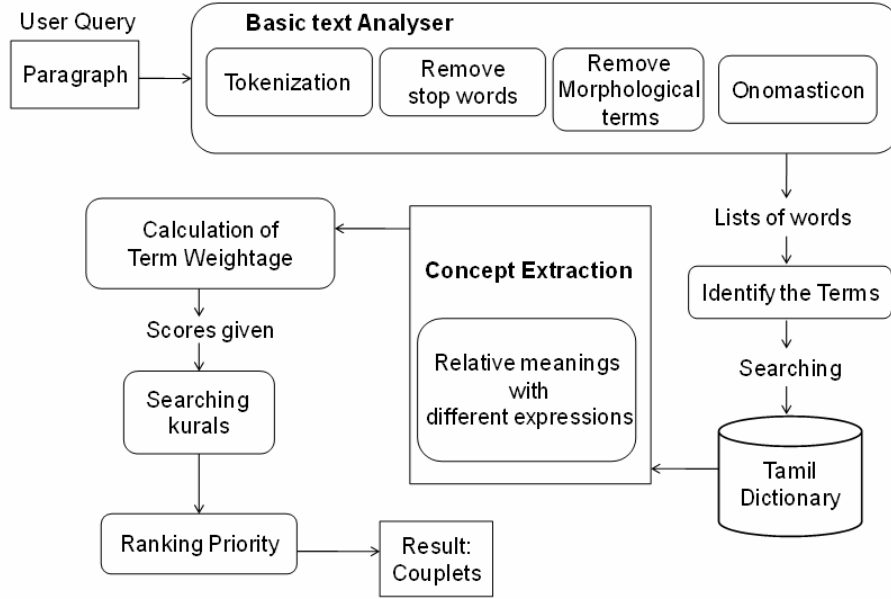
The context based Information search can be obtained by identifying the relativity among terms and derive a context from the terms. The term ranking can be calculated using Term frequency in the given sentence or paragraph of words.

Users query could be answered in semantic approach and get the couplets in ranking order. The above architecture diagram clearly explains the flow of the paper.

The system gets the input text as sentences like paragraph for searching information. The given sentence or paragraph is tokenized and each token is passed to Tamil language analyzer. The analyzer identifies the part of the syntax such as noun, verb, adjective, pronoun, adverb and so on that the token belongs to. The analyzer removes the common Tamil morphological. The system uses the noun, verb, adjective and adverbs to find the relation among the terms.

The relativity among the terms can be derived using a Tamil dictionary. The term weightage is calculated for each relative term using the term weightage formula. The context is extracted from the relative terms and finally searches the couplets in a Thirukural database.

Figure: Overall Architecture



Term Weightage Formula

By using Term Weightage Formula, calculates the weightage of the words which could be in any form as related meaning, different expression and comparative meaning. In a paragraph some of the sentences have same meaning but in different expressions by using term weightage formula, a well formed questions term having the weightage.

$$t_w = nt_r / T$$

where as,

t_w - term weightage

nt_r - number of related terms.

T- Total number of terms identified in the given sentence or paragraph.

Calculate the total number of terms and the related meaning and the terms which have relative meaning.

For example, the following sentence shown as,

உலகத்தில் வாழும் அனைத்து மக்களுக்கும் செல்வம் நிறைந்து இருப்பதற்கும் போற்றுவதற்கும் கல்வி கட்டாயம் தேவை.

In the given sentence, the terms are analyzed and grouped into as proper names of all kinds and the origins of names.

List of terms :

- உலகம்: Noun
- வாழ்: Verb
- அணைத்து: Adverb
- மக்கள்: Noun
- செல்வம்: Noun
- நிறை: Verb
- இருப்பு: Noun
- போற்று: Verb
- கல்வி: Noun
- கட்டாயம்: Adverb
- தேவை: Verb

[உலகம், வாழ்] [தேவை] [இருப்பு] [நிறை] [போற்று] [கல்வி] [செல்வம்] [மக்கள்]

Each term have their respective weightage and the related terms have more weightage than the remaining terms

[வாழ், உலகம்] is the relative meaning of the above example and this could be calculated as

$$\begin{array}{ccc} & \text{[வாழ், உலகம்]} & \\ & \downarrow & \\ \left. \begin{array}{l} \text{[வாழ்]=0.222,} \\ \text{கல்வி=0.111,} \\ \text{நிறை=0.111,} \\ \text{தேவை=0.111,} \\ \text{உலகம்=0.222,} \end{array} \right\} & \frac{2}{11} = 0.222 & \left. \begin{array}{l} \text{செல்வம்=0.111,} \\ \text{இருப்பு=0.111,} \\ \text{போற்று=0.111,} \\ \text{மக்கள்=0.111} \end{array} \right\} \end{array}$$

Result as,

1. நிறையுடைமை நீங்காமை வேண்டின் பொறையுடைமை போற்றி யொழுகப் படும்.
2. கெடுவாக வையா துலகம் நடுவாக நன்றிக்கண் தங்கியான் தாழ்வு.
3. எவ்வ துறைவ துலக முலகத்தோ டவ்வ துறைவ தறிவு.

Conclusion

Thirukkural is a treatise par excellence on the art of living and the real greatness of Thirukkural is its survival, even after the onslaughts of many heterogenous creeds. Each and everything in our life is related to the theme of Thirukural. By giving a paragraph the search engine analyse the context of the sentences and groups the related terms to find out the equivalent meanings of couplets in ranking order with the help of term weightage. This search engine helps the users to easily gather the information of couplets for any paragraph in the context based.

References

1. Akshay Java, Sergei Nirunburg, Timothy Finin, Jesse English, Anupam Joshi, "Using a Natural Language Understanding System to Generate Semantic Web Content" International Journal on Semantic web and Information system, 2007.
2. Fuji Ren and Tianjiao Gu, "Question Matching based on Fuzzy Set", Faculty of Engg, The University of Tokushima, IEEE International conference, 2008
3. Jamie Callan, "Human Language Technologies, Open Domain Question Answering" Carnegie Mellon University, 2004.
4. Jun'ichi FUKUMOTO, Tsuneaki KATO, "An Overview of Question and Answering Challenge (QAC)" Ritsumeikan University IEEE, 2002.
5. Mitsuru Ishizuka "A Common Concept Description of Natural Language Texts as the Foundation of Semantic Computing on the Web" The University of Tokyo.
6. Natural Language Processing, "<http://research.microsoft.com/en-us/groups/nlp>".
7. Thorsten Brants, "Natural Language Processing in Information Retrieval", 2003.
8. "Tamil Virtual University" enable Tamil Education easily and effectively.
9. "<http://www.tamilvu.org>".

அசைநிலை உரை வடிவப் பாக்களுக்கான இடஞ்சாரா இலக்கணமும் பகுப்பாய்வு முறையும்

A Context-free grammar and a method to parse verses in metrical text

இல.பாலசுந்தரராமன், ஈசுவர் சிரீதரன்

sundarbese@yahoo.com, ishwar@gmail.com

கட்டுரைச் சுருக்கம்

எழுத்தசை சீர்தளை அடிதொடை தூக்கொடு
இழுக்கா நடைய தியாப்பெனப் படுமே.

யாப்பிலக்கண நூல்கள் தமிழ் மரபுப்பாக்களை எழுத்து, அசை, சீர், தளை, அடி, தொடை என்ற உறுப்புகளைக் கொண்டு நான்கு பா வகைகளாகவும் அவற்றின் உள்வகைகளையும் சேர்த்து 12 பாவினங்களாகவும் வகுத்துள்ளன. நால்வகைப் பாக்களுக்கான இடஞ்சாரா இலக்கணத்தை இக்கட்டுரை விவரிக்கிறது.

பழந்தமிழ் நூல்களின் பா வடிவங்களும், புணர்ச்சி பிரிக்கப்பட்ட வடிவங்களும் புழக்கத்தில் உள்ளன. இவ் இருமை நிலையைப் பற்றி ஏற்கனவே ஆய்வுக் கட்டுரை ஒன்றில் விளக்கப்பட்டுள்ளது. மேலும், மரபின் பொருட்டும், இசை நயத்துக்கெனவும் நினைவுக்குறிப்புக்காகவும் இரு வடிவங்களையும் பாதுகாக்க வேண்டியது இன்றியமையாதது.

இத்தேவையை முன்னிட்டு மேலே குறிப்பிட்டுள்ள இடஞ்சாரா இலக்கணத்தின் அடிப்படையில் நாங்கள் உருவாக்கியுள்ள விசைநெறி என்ற பகுப்பாய்வியையும் (parser) இக்கட்டுரையில் அறிமுகப்படுத்துகிறோம். இப்பகுப்பாய்வியின் நீட்டவல்ல கட்டமைப்பை இக்கட்டுரையில் விளக்குகிறோம். இது பா வடிவ உரையை அலகு பிரித்து அதன் கூறுகளைப் பதிவு செய்வதற்காக முன்னர் முன்வைக்கப்பட்ட குறியீட்டு முறையில் (markup scheme) தனது வெளியீட்டை சேமிக்கிறது.

யாப்பிலக்கண உறுப்புகளும் பாவகைகளும்

எழுத்து, அசை, சீர், தளை, அடி, தொடை ஆகியன யாப்பிலக்கண உறுப்புகளாக அறியப்படுவன. இவற்றுள் சில தொடைகளை விடுத்தால் மற்றவை ஒலிப்பியல் அளவில் பாக்களை நெறிப்படுத்தும் உறுப்புகள் ஆவன. இவ்வுறுப்புகளைக் கொண்டு விளங்கும் யாப்பிலக்கண நெறிகளின்படி அமைந்த பாக்கள் நான்கு வகைகளாகவும் பல்வேறு இனங்களாகவும் வகுக்கப்பட்டுள்ளன.^[1]

யாப்பிலக்கணத்தில் எழுத்து குறில், நெடில், ஒற்று, குற்றியலிகரம், அளபெடை, குற்றியலுகரம் என வகைப்படும். நேர், நிரை என அசைகள் இரண்டு. தாம் கொண்டிருக்கும் அசைகளின் எண்ணிக்கையைப் பொருத்து சீர்கள் இயற்சீர், உரிச்சீர், பொதுச்சீர் என்று வகைப்படுத்தப்படும். ஒன்றிலிருந்து நான்கு அசைகள் வரை பெற்று வரும் சீர்கள் உள்ளன.

^[1] <https://www.tamilvu.org/courses/diploma/a021/a0214/html/a0214410.htm>

ஒரு சீரின் வகைக்கும் அடுத்து வரும் சீரின் முதலசைக்கும் உள்ள தொடர்பைப் பொருத்து தளைகள் நேரொன்றாசிரியத்தளை, நிரையொன்றாசிரியத்தளை, இயற்சீர் வெண்டளை, வெண்சீர் வெண்டளை, கலித்தளை, ஒன்றிய வஞ்சித்தளை, ஒன்றாத வஞ்சித்தளை என வகைப்படும். சீர்களின் எண்ணிக்கையின் அடிப்படையில் அடிகளை குறளடி, சிந்தடி, அளவடி, நெடிவடி, கழிநெடிவடி என வகைப்படுத்துவர்.

ஒரு பாவில் உள்ள வெவ்வேறு அடிகளின் இடையேயும் ஒரு அடியின் சீர்களின் இடையேயும் அமையும் தொடர்பு தொடையாகும். மோனை, எதுகை, முரண், இயைபு, அளபெடை, அந்தாதி, இரட்டை, செந்தொடை என்ற தொடை வகைகளுள் முரணையும் செந்தொடையையும் நீக்கினால் எஞ்சியவை அனைத்தும் ஒலிப்பியல் அடிப்படையில் அமைந்துள்ளபடியால் எளிதில் பிரித்துணர இயலும்.

இடஞ்சாரா இலக்கணம்

மொழியியலில் சொற்றொடர் அமைப்பு இலக்கணம் என்றும் அறியப்படும் இடஞ்சாரா இலக்கணம் என்பது முறைப்படியான கட்டுக்கோப்பான மொழிப் பகுதிகளுக்கான இலக்கணமாகும். இவை பல அடுக்குகளாக உறுப்புகளைப் பிரித்துப் படிநிலை போன்று அமைக்கக் கூடிய இலக்கணங்களாகும், ஆனால் ஒரு உறுப்பும் மற்றொன்றும் ஒன்றன் மேல் ஒன்று படிதல் கூடாது.

இவ்விலக்கண முறையை நோம் சாம்சுக்கி 1956-ல் இயல் மொழிகளின் இலக்கணங்களைக் குறிக்கும் நோக்கில் அறிமுகம் செய்தார்.^[2] சுருங்குறித்தொடர்களைக் கொண்டு இயற்றப்படும் சீருற்ற இலக்கணங்களைக் (regular grammars) காட்டிலும் இடம் சாரா இலக்கணங்கள் பகர்திறன் மிகுதியாகக் கொண்டவை என்று அவர் நிறுவினார். ஆனால் எந்த ஒரு இயல் மொழியின் இலக்கணத்தையும் முழுமையாக இடம் சாரா இலக்கணத்தைக் கொண்டு வரையறுக்க இயலுமென அவரால் உறுதிபடக் காட்ட முடியவில்லை. இன்றைய ஆய்வர்கள் இயல்மொழி இலக்கணங்கள் இடம் சாராதவை அல்ல என்றே இணங்கியுள்ளனர்.^[3] அதே வேளையில் இயல்மொழிகளின் பெரும்பகுதி இடம்சாரா இலக்கணம் கொண்டது என்று காட்டியுள்ளனர்.^[4] இது போன்ற பகுதிகளுக்கான இடம் சாரா இலக்கணங்களையும் எழுதியுள்ளனர்.^{[5][6]}

பேக்கசு-நார் முறை

கணினி மென்பொருள்கள் உணருவதற்கு ஏற்ற வகையிலும், இலக்கணவியலாளர்கள் படித்துப் புரிந்து கொள்ளத் தக்க வகையிலும் இடஞ்சாரா இலக்கணத்தைக் குறிப்பதற்கு உதவும் முறை பேக்கசு-நார் முறையாகும். சா'ன் பேக்கசும் பீட்டர் நாரும் இம்முறையை உருவாக்கினர்.^[7]

செய்யுள் இருமை நிலையும் அசைநிலை வடிவத்தின் தேவையும்

பழந்தமிழ் நூல்களின் பா வடிவங்களும், புணர்ச்சி பிரிக்கப்பட்ட வடிவங்களும் புழக்கத்தில் உள்ளன. இவ் இருமை நிலையைப் பற்றி முன்னதாக இழான்-லூய்க் செவ்வியார் விளக்கியுள்ளார்.^[8] தெளிவுரை நூல்கள் பொருள் காட்டும் பொருட்டு புணர்ச்சி பிரித்துக் காட்டுகின்றன. ஓதுவரும், பாக்களைப் பாடுபவர்களும் பா வடிவங்களை உள்ளபடியே படித்துப் பாடுகின்றனர். மரபின் பொருட்டும், இசை நயத்துக்கெனவும் நினைவுக்குறிப்புக்காகவும் இரு வடிவங்களையும் பாதுகாக்க வேண்டியது இன்றியமையாதது.^{[8][9]}

தேவாரப் பாடல் ஒன்றின் இரு வடிவங்களைக் காண்க:

பா வடிவம்^[10]

புணர்ச்சி பிரித்த வடிவம்^[11]

தோடுடைய செவியன் விடையேறியோர் தூவெண் மதிசூடிக்
காடுடைய சுடலைப் பொடிபூசியென் னுள்ளங் கவர்கள்வன்
ஏடுடைய மலரான் முனைநாட்பணிந் தேத்த வருள்செய்த
பீடுடைய பிரமா புரமேவிய பெம்மா னிவனன்றே.

தோடு உடைய செவியன், விடை ஏறி, ஓர் தூ வெண்மதி சூடி,
காடு உடைய சுடலைப் பொடி பூசி, என் உள்ளம் கவர் கள்வன்---
ஏடு உடைய மலரான் முனைநாள் பணிந்து ஏத்த, அருள்செய்த,
பீடு உடைய பிரமாபுரம் மேவிய பெம்மான்---இவன் அன்றே!

யாப்பிலக்கண உருவகங்கள்

நால்வகைப் பாக்கள் தாம் பெற்று வரும் சீர்களைப் பொருத்தும், தளைகளைப் பொருத்தும் தம் இலக்கணத்தில் மாறுபடும். இவற்றுக்கான பொதுவான இலக்கண உருவகங்களைக் கீழே காணலாம். பின்வரும் இலக்கண உருவகங்கள் பகுப்பாய்வி வடிக்கும் மென்பொருளின் தேவைக்கேற்ப நீட்டிய பேக்கசு-நார் முறையில் (EBNF) எழுதப்பட்டுள்ளன. ஒவ்வொரு பா வகையும் கூடுதலாகச் சில சிறப்பு உருவகங்களைப் பெற்றிருக்கும். அவை இங்கே தரப்படவில்லை. மேலும் சில நெறிமுறைகள் எளிமைப்படுத்தப்பட்டுள்ளன. மொழியிலாளர்களின் பின்னூட்டத்துடன் இவை மேம்படுத்தப்படும்.

பா ::= அடி பா

(யாப். 6,7)

கருவிளம் ::= நிரை நிரை

பா ::= அடி

நேர் ::= குறில் ஒற்று

புளிமா ::= நிரை நேர்

நேர் ::= நெடில் ஒற்று

கூவிளம் ::= நேர் நிரை

[2] Chomsky, Noam (1956). "Three models for the description of language". IRE Transactions on Information Theory (2): 113-124.

[3] Shieber, Stuart (1985). "Evidence against the context-freeness of natural language". Linguistics and Philosophy 8: 333-343. doi:10.1007/BF00630917.

[4] Pullum, Geoffrey K.; Gerald Gazdar (1982). "Natural languages and context-free languages". Linguistics and Philosophy 4: 471-504. doi:10.1007/BF00360802

[5] L, BalaSundaraRaman; Ishwar.S, Sanjeeth Kumar Ravindranath (2003-08-22). "Context Free Grammar for Natural Language Constructs - An implementation for Venpa Class of Tamil Poetry". Proceedings of Tamil Internet, Chennai, 2003. International Forum for Information Technology in Tamil. pp. 128-136.

[6] M.G.J. van den Brand, M.P.A. Sellink, and C. Verhoef (2000). "Generation of Components for Software Renovation Factories from Context-free Grammars". Science of Computer Programming 36: 209-266.

[7] Knuth, Donald E. (1964). "Backus Normal Form vs. Backus Naur Form". Communications of the ACM 7 (12): 735-736. doi:10.1145/355588.365140.

[8] Jean-Luc Chevillard, Critical editions of Tamil works: exploratory survey and future perspectives (INFITT 2009, Köln, 25th October)

[9] W. Van Peer (1990) The measurement of metre : Its cognitive and affective functions. Poetics 19 (3): 259 - 275.

[10] http://www.shaivam.org/tamil/thirumurai/thiru01_001.htm

[11] http://www.ifpindia.org/ecrre/upload/digital_database/Site/Digital_Tevaram/U_TEV/DM1_1.HTM

அடி ::= சீர் இடைவெளி அடி
 அடி ::= சீர் அடிமுடிபு
 அடி ::= ஈற்றுச்சீர் அடிமுடிபு

சீர் ::= ஈரசைச்சீர்
 சீர் ::= மூவசைச்சீர்
 சீர் ::= நாலசைச்சீர்

ஈற்றுச்சீர் ::= ஓரசைச்சீர்
 ஈற்றுச்சீர் ::= நிரைபு
 ஈற்றுச்சீர் ::= நேர்பு

(யாபு. 14)

ஓரசைச்சீர் ::= மலர்
 ஓரசைச்சீர் ::= நாள்

மலர் ::= குறில் குறில் ஒற்று
 மலர் ::= குறில் கடைக்குறில்
 மலர் ::= குறில் நெடில் ஒற்று
 மலர் ::= குறில் கடைநெடில்

(யாபு. 11)

ஈரசைச்சீர் ::= கருவிளம்
 ஈரசைச்சீர் ::= புளிமா
 ஈரசைச்சீர் ::= கூவிளம்
 ஈரசைச்சீர் ::= தேமா

(யாபு. 12)

மூவசைச்சீர் ::= கருவிளங்கனி
 மூவசைச்சீர் ::= கருவிளங்காய்
 மூவசைச்சீர் ::= புளிமாங்கனி
 மூவசைச்சீர் ::= புளிமாங்காய்
 மூவசைச்சீர் ::= கூவிளங்கனி
 மூவசைச்சீர் ::= கூவிளங்காய்
 மூவசைச்சீர் ::= தேமாங்கனி
 மூவசைச்சீர் ::= தேமாங்காய்

நேர் ::= கடைக்குறில்
 நேர் ::= கடைநெடில்
 நேர் ::= குறில்

(யாபு. 8,9)

நிரை ::= குறில் குறில் ஒற்று
 நிரை ::= குறில் கடைக்குறில்
 நிரை ::= குறில் குறில்
 நிரை ::= குறில் நெடில் ஒற்று
 நிரை ::= குறில் கடைநெடில்
 நிரை ::= குறில் நெடில்

நேர்பு# ::= நேர் குற்றியலுகரம்
 நிரைபு# ::= நிரை குற்றியலுகரம்
 # எளிமைப்படுத்திய நெறிகள்

(யாபு. 13)

நாலசைச்சீர் ::= கருவிளநறுநிழல்
 நாலசைச்சீர் ::= கருவிளநறும்பூ
 நாலசைச்சீர் ::= கருவிளந்தண்ணிழல்
 நாலசைச்சீர் ::= கருவிளந்தண்பூ
 நாலசைச்சீர் ::= புளிமாநறுநிழல்
 நாலசைச்சீர் ::= புளிமாநறும்பூ
 நாலசைச்சீர் :: புளிமாந்தண்ணிழல்
 நாலசைச்சீர் ::= புளிமாந்தண்பூ
 நாலசைச்சீர் ::= கூவிளநறுநிழல்
 நாலசைச்சீர் ::= கூவிளநறும்பூ
 நாலசைச்சீர் ::= கூவிளந்தண்ணிழல்
 நாலசைச்சீர் ::= கூவிளந்தண்பூ
 நாலசைச்சீர் ::= தேமாநறுநிழல்
 நாலசைச்சீர் ::= தேமாநறும்பூ
 நாலசைச்சீர் ::= தேமாந்தண்ணிழல்
 நாலசைச்சீர் ::= தேமாந்தண்பூ

தேமா ::= நேர் நேர்

கருவிளங்கனி ::= நிரை நிரை நிரை
 புளிமாங்கனி ::= நிரை நேர் நிரை
 கூவிளங்கனி ::= நேர் நிரை நிரை
 தேமாங்கனி ::= நேர் நேர் நிரை
 கருவிளங்காய் ::= நிரை நிரை நேர்
 புளிமாங்காய் ::= நிரை நேர் நேர்
 கூவிளங்காய் ::= நேர் நிரை நேர்
 தேமாங்காய் ::= நேர் நேர் நேர்

கருவிளநறுநிழல் ::= நிரை நிரை நிரை நிரை
 கூவிளநறுநிழல் ::= நேர் நிரை நிரை நிரை
 புளிமாநறுநிழல் ::= நிரை நேர் நிரை நிரை
 தேமாநறுநிழல் ::= நேர் நேர் நிரை நிரை
 கருவிளநறும்பூ ::= நிரை நிரை நிரை நேர்
 கூவிளநறும்பூ ::= நேர் நிரை நிரை நேர்
 புளிமாநறும்பூ ::= நிரை நேர் நிரை நேர்
 தேமாநறும்பூ ::= நேர் நேர் நிரை நேர்
 கருவிளந்தண்ணிழல் ::= நிரை நிரை நேர் நிரை
 கூவிளந்தண்ணிழல் ::= நேர் நிரை நேர் நிரை
 புளிமாந்தண்ணிழல் ::= நிரை நேர் நேர் நிரை
 தேமாந்தண்ணிழல் ::= நேர் நேர் நேர் நிரை
 கருவிளந்தண்பூ ::= நிரை நிரை நேர் நேர்
 கூவிளந்தண்பூ ::= நேர் நிரை நேர் நேர்
 புளிமாந்தண்பூ ::= நிரை நேர் நேர் நேர்
 தேமாந்தண்பூ ::= நேர் நேர் நேர் நேர்

இலக்கண பகுப்பாய்வுப் படிநிலை வெளியீட்டு முறை

பாக்களின் இருவடிவங்களையும் அடக்கி, அலகிட்டு, யாப்பின் உறுப்புகளைக் குறிப்பிட்டு ஆவணப்படுத்துவதற்கான வெளியீட்டு முறை இன்றியமையாதது. உரைகளுக்கான குறிமுறை முனைப்பின் (Text Encoding Initiative) P5 நெறிமுறைகள் பன்மொழி உரைநடை மற்றும் பண்களின் இலக்கண மரபுகளை நீட்டவல்ல குறியீட்டு முறையில்(XML) வெளிப்படுத்துவதற்கான வெளியீட்டு முறை ஒன்றைப் பரிந்துரைத்துள்ளது.^[12] இந்தப் பரிந்துரையின் அடிப்படையில் தமிழ் பாக்களுக்கான வெளியீட்டு முறையொன்றை இழான்-லூய்க் செவ்வியார் பரிந்துரைத்துள்ளார். ^[8] அவரது பரிந்துரையை ஏற்று, விசைநெறியின் வெளியீட்டு முறையாக, எஞ்சிய தகவல்களான தளை, அடி வகை போன்ற தகவல்களையும் அதே கட்டமைப்பில் இணைத்துத் தரும்படி அமைத்துள்ளோம். பா, அடி, சீர் போன்ற உறுப்புகளை XML கணுக்களாகக்^[13] கொண்டு இவற்றுக்கிடையே அமைந்த உறவுகளான தளை, தொடை போன்றவற்றை அக்கணுக்கள் மீதான பண்படைகளாய் (attribute) ஏற்றி வெளியிடுகிறோம்.

WE USE TEI

எடுத்துக்காட்டு

தோடுடைய செவியன் விடையேறியோர்
துாவெண் மதிசூடிக்க

காடுடைய சுடலைப் பொடிபூசியென் னுள்ளங்
கவர்கள்வன்

ஏடுடைய மலரான் முனைநாட்பணிந் தேத்த
வருள்செய்த

பீடுடைய பிரமா புரமேவிய பெம்மா
னிவனன்றே.

```

- <பாவரிசை>
- <பா>
- <அடி பொருள்="தோடுடைய செவியன் விடையேறியோர் துாவெண் மதிசூடிக்க">
  <சீர் இனம்="கவிளங்காய்">தோடுடைய</சீர்>
  <சீர் இனம்="புளிமா">செவியன்</சீர்>
  <சீர் இனம்="புளிமாங்கனி">விடையேறியோர்</சீர்>
  <சீர் இனம்="தேமா">துாவெண்</சீர்>
  <சீர் இனம்="புளிமாங்காய்">மதிசூடிக்க</சீர்>
</அடி>
- <அடி பொருள்="காடுடைய சுடலைப் பொடிபூசியென் னுள்ளங் கவர்கள்வன்">
  <சீர் இனம்="கவிளங்காய்">காடுடைய</சீர்>
  <சீர் இனம்="புளிமா">சுடலைப்</சீர்>
  <சீர் இனம்="புளிமாங்கனி">பொடிபூசியென்</சீர்>
  <சீர் இனம்="தேமா">னுள்ளங்</சீர்>
  <சீர் இனம்="புளிமாங்காய்">கவர்கள்வன்</சீர்>
</அடி>
- <அடி பொருள்="ஏடுடைய மலரான் முனைநாட்பணிந் தேத்த வருள்செய்த">
  <சீர் இனம்="கவிளங்காய்">ஏடுடைய</சீர்>
  <சீர் இனம்="புளிமா">மலரான்</சீர்>
  <சீர் இனம்="புளிமாங்கனி">முனைநாட்பணிந்</சீர்>
  <சீர் இனம்="தேமா">தேத்த</சீர்>
  <சீர் இனம்="புளிமாங்காய்">வருள்செய்த</சீர்>
</அடி>
- <அடி பொருள்="பீடுடைய பிரமா புரமேவிய பெம்மா நிவனன்றே.">
  <சீர் இனம்="கவிளங்காய்">பீடுடைய</சீர்>
  <சீர் இனம்="புளிமா">பிரமா</சீர்>
  <சீர் இனம்="புளிமாங்கனி">புரமேவிய</சீர்>
  <சீர் இனம்="தேமா">பெம்மா</சீர்>
  <சீர் இனம்="புளிமாங்காய்">னிவனன்றே.</சீர்>
</அடி>
</பா>
</பாவரிசை>

```

இங்கு “பொருள்” என்ற பண்படையில் புணர்ச்சி பிரித்த வடிவை இட்டுவிட்டால் பாக்களை எவ்வடிவிலும் தேட ஏதுவாகும். அதற்கான வசதியை இம்மென்பொருளில் எளிதில் சேர்க்க முடியும். மேலும், தளை போன்ற பண்புகள் சில பா வகைகளுக்கே பொருந்தும் என்றாலும் ஆய்வு நோக்கில் இங்கு பகுத்துக் காட்டியுள்ளோம்.

^[12] TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [November 1, 2007]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>

^[13] <http://ta.wiktionary.org/wiki/node>

பகுப்பாய்வியின் கட்டமைப்பு

முன்னர் காட்டியுள்ள யாப்பிலக்கணத்துக்கான இடஞ்சாரா இலக்கணத்தின் அடிப்படையில் நாங்கள் உருவாக்கியுள்ள மென்பொருள் "விசைநெறி" ஆகும். உள்ளிடப்படும் ஒரு பா, பேக்கசு-நார் குறியீட்டு முறையில் அமைந்துள்ள இவ்விலக்கண நியமங்களுக்கு ஏற்ப அமைந்துள்ளதா எனப் பகுப்பாய்ந்து, அவ்வாறிருப்பின் அப்பாவின் எழுத்து, அசை, சீர், தளை, அடி, தொடை போன்ற உறுப்புக்களை அடையாளம் கண்டு மேலே விளக்கியுள்ள XML குறியீட்டு முறையில் இப்பகுப்பாய்வி வெளியிடுகிறது.

விசைநெறி, பைத்தான் நிரல்மொழியில் உள்ள spark^[14] எனும் பகுப்பாய்விக் கட்டமைப்பைப் பற்றி அமைந்துள்ளது. இக்கட்டமைப்பின்படி, இப்பகுப்பாய்வி கண்ணி, பகுப்பி, பொருண்மை அலசி^[15] என மூன்று கட்டங்களாக அமைக்கப்பட்டுள்ளது.

1. சொன்மைக் கண்ணி (lexical scanner)

இக்கட்டத்தின்போது, பாவின் எழுத்துக்கள் குறில், நெடில், ஒற்று, குற்றியலுகரம், குற்றியலிகரம் என வகைபிரிக்கப்பட்டு குறியிடப்படும்.

2. தொடரொழுக்கியல் அலசி/பகுப்பி (syntactic analyser)

பாவிலக்கணப்படி குறில், நெடில் மற்றும் ஒற்றுக்களின் இடையேயுள்ள உறவை விளக்குவது அசை எனும் யாப்பிலக்கண உறுப்பாகும். அதேபோல், அசைகளின் இடையேயுள்ள உறவை விளக்குவது சீர் ஆகும். பல்வேறு சீர்களைக் கொண்டு அடி அமைக்கப்படும்.

இக்கட்டத்தில், குறில், நெடில் மற்றும் ஒற்றாக அடையாளங் காணப்பட்ட குறிகளைக்கொண்டு அசை, சீர் மற்றும் அடிகள் பிரிக்கப்பட்டு குறியிடப்படும்.

3. பொருட்குறிப்பியல் அலசி (semantic analyser)

தளை, தொடை, அணி எனும் யாப்பிலக்கண உறுப்புக்கள் சீர்களினிடையேயுள்ள உறவினைக் குறிக்கும். இக்கட்டத்தில், முன்னர் அசை, சீர் மற்றும் சீர்களாக பிரிக்கப்பட்ட பாவைக்கொண்டு தளை, தொடை, அணி எனும் உறவுகள் கண்டுபிடிக்கப்பட்டு குறியிடப்படும்.

எடுத்துக்காட்டு:

தோடுடைய செவியன் விடையேறியேயார் தூவெண் மதிசூடிக்

- கண்ணி வெளியீடு:

தோ:நெடில், டு:குறில், டை:குறில், ய:குறில் செ:குறில், வி:குறில், ய:குறில், ன்:ஒற்று வி:குறில், டை:குறில், யே:நெடில், றி:குறில், யோ:நெடில், ர்:ஒற்று தூ:நெடில், வெ:குறில், ண்:ஒற்று ம:குறில், தி:குறில், சூ:நெடில், டி:குறில், க்:ஒற்று

- பகுப்பி படிநிலை வெளியீடு:

<பா>

<அடி பொருள்="தோடுடைய செவியன் விடையேறியேயார் தூவெண் மதிசூடிக்">

<சீர் இனம்="கூவிளங்காய்">தோடுடைய</சீர்>

<சீர் இனம்="புளிமா">செவியன்</சீர்>

^[14] <http://pages.cpsc.ucalgary.ca/~aycock/spark/>

^[15] http://groups.google.com/group/tamil_wiktionary/msg/30af374714abb5ba?

<சீர் இனம்="புளிமாங்கனி">விடையேறியோர்</சீர்>
 <சீர் இனம்="தேமா">தூவெண்</சீர்>
 <சீர் இனம்="புளிமாங்காய்">மதிசூடிக்க</சீர்>
 </அடி>
 o பொருண்மை அலசி வெளியீடு:
 <பா>
 <அடி பொருள்="தோடுடைய செவியன் விடையேறியோர் தூவெண் மதிசூடிக்க" வகை="நெடிலடி">
 <சீர் இனம்="கூவிளங்காய்" தளை="கலித்தளை">தோடுடைய</சீர்>
 <சீர் இனம்="புளிமா" தளை="இயற்சீர் வெண்டளை">செவியன்</சீர்>
 <சீர் இனம்="புளிமாங்கனி" தளை="ஒன்றாத வஞ்சித்தளை">விடையேறியோர்</சீர்>
 <சீர் இனம்="தேமா" தளை="இயற்சீர் வெண்டளை">தூவெண்</சீர்>
 <சீர் இனம்="புளிமாங்காய்" தளை="வெண்சீர் வெண்டளை">மதிசூடிக்க</சீர்>
 </அடி>

பகுப்பாய்வியின் யன்கள்

இந்தப் பா அலகீட்டு மென்பொருள் விசைநெறியின் பயன்கள் பின்வருமாறு:

1. யாப்பிலக்கணம் கட்டுக்கோப்பான பகுப்பாய்வுக்கேற்ற இடஞ்சாரா இலக்கணம் என நிறுவுதல்
2. தமிழ் இலக்கியங்களைப் பகுப்பாய்ந்து சீர் வகைகள், தளைகள், பாவினங்கள் ஆகியவை தொடர்பான தகவல் புள்ளிகளைத் தொகுக்க வகை செய்தல்
3. இந்நாள் கவிஞர்கள் பா எழுதத் துணை செய்தல்^[16]
4. இதன் பகுதிகளான இலத்தி (tokeniser) போன்றவை பிற இயல்மொழி ஆய்வுகளுக்கு உதவுதல்

முடிவுரை

மரபுப்பாவினங்களுக்கான யாப்பிலக்கணத்தை இடஞ்சாரா இலக்கணமாக நீட்டிய பேக்கசு-நார் முறையில் எழுத முடியும் என்றும் அவ்வாறான பாக்களை மென்பொருள் கொண்டு பகுக்கவியலும் என்றும் நிறுவப்பட்டுள்ளது. இத்தகைய மென்பொருளான விசைநெறியின் கட்டமைப்பை விளக்கி, அதன் வெளியீட்டு முறையாக ஒரு நீட்டவல்ல குறியீட்டுமுறையையும் பரிந்துரைத்துள்ளோம்.

உசாத்துணை

* யாப்பருங்கலம்

மேற்கோள்கள்

- [1] <https://www.tamilvu.org/courses/diploma/a021/a0214/html/a0214410.htm>
- [2] Chomsky, Noam (1956). "Three models for the description of language". IRE Transactions on Information Theory (2): 113–124.
- [3] Shieber, Stuart (1985). "Evidence against the context-freeness of natural language". Linguistics and Philosophy 8: 333–343. doi:10.1007/BF00630917.
- [4] Pullum, Geoffrey K.; Gerald Gazdar (1982). "Natural languages and context-free languages". Linguistics and Philosophy 4: 471–504. doi:10.1007/BF00360802

^[16] http://groups.google.com/group/anbudan/browse_thread/thread/fc8c416a2b42f73a/00efb259006416d1?pli=1

- [5] L. BalaSundaraRaman; Ishwar.S, Sanjeeth Kumar Ravindranath (2003-08-22). "Context Free Grammar for Natural Language Constructs - An implementation for Venpa Class of Tamil Poetry". Proceedings of Tamil Internet, Chennai, 2003. International Forum for Information Technology in Tamil. pp. 128-136.
- [6] M.G.J. van den Brand, M.P.A. Sellink, and C. Verhoef (2000). "Generation of Components for Software Renovation Factories from Context-free Grammars". Science of Computer Programming 36: 209-266.
- [7] Knuth, Donald E. (1964). "Backus Normal Form vs. Backus Naur Form". Communications of the ACM 7 (12): 735-736. doi:10.1145/355588.365140.
- [8] Jean-Luc Chevillard, Critical editions of Tamil works: exploratory survey and future perspectives (INFITT 2009, Köln, 25th October)
- [9] W. Van Peer (1990)The measurement of metre : Its cognitive and affective functions. Poetics 19 (3): 259 - 275.
- [10] http://www.shaivam.org/tamil/thirumurai/thiru01_001.htm
- [11] http://www.ifpindia.org/ecrere/upload/digital_database/Site/Digital_Tevaram/U_TEV/DM1_1.HTM
- [12] TEI Consortium, eds. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [November 1, 2007]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>
- [13] <http://ta.wiktionary.org/wiki/node>
- [14] <http://pages.cpsc.ucalgary.ca/~aycock/spark/>
- [15] http://groups.google.com/group/tamil_wiktioary/msg/30af374714abb5ba?
- [16] http://groups.google.com/group/anbudan/browse_thread/thread/fc8c416a2b42f73a/00efb259006416d1?pli=1

Computational approaches for learning inflections in Tamil

K.Rajan, Dept of Comp. Sc. and Engineering

V.Ramalingam, Dept of Comp Sc. and Engineering

M.Ganesan, Centre of Advanced Studies in Linguistics

Annamalai University, Annamalainagar

Abstract

This paper proposes machine learning techniques for the study of inflections in Tamil. In recent years there has been growing interest in using machine learning techniques for natural language processing tasks. Recent computational research on natural language corpora has revealed that simple machine learning mechanisms could make an important contribution to certain aspects of language acquisition. There are regularities that can be captured mechanically, by exploiting distributional patterns found in the language data. Very large corpora will yield to mathematical model of language systems.

The Artificial neural network (ANN) is used to learn the regularities of Tamil morphology. In this paper, the methods of feature extraction and feature representation for Tamil verb morphology are discussed. The ANN network with three layers is trained with the features extracted from different verb conjugations generated by the morphological generator. More than 7800 word forms are used for training and testing this model. The network produces the grammatical values as output for the morphemes present in the given word. This model does not require morphological rules. The output is comparable with the results of rule based morphological analyser. The performance of this model on different verb types is presented.

Keywords: Machine learning, Tamil morphology, ANN Language Model

Introduction

In recent years there has been growing interest in using machine learning techniques for natural language processing tasks. In this paper, the Artificial neural network is trained to learn the relationship between the morphemes and their categories from the training set of inflected Tamil verb forms. Morphemes are portions of a word that recur in other words with the same meaning. They are minimal, that they cannot be broken into pieces. The grammatical value of the whole word is related to the component morphemes of the word. In highly-inflecting and compounding languages the number of possible word forms is very high. This poses special challenges to Natural Language Processing systems dealing with these languages.

The morphological tagging is an important problem in the area of computational linguistics, as it underlies other crucial tasks such as syntactic parsing and machine translation. Artificial neural network is a promising approach to this kind of problem, for which the exact algorithmic solution is unknown or

not efficient enough. In this paper we present the results obtained by the application of neural network model trained with back propagation algorithm for Tamil morphology.

Knowledge discovery in text is a non-trivial process of identifying valid, novel, potentially useful, and understandable patterns in unstructured text data. Learning algorithms are an integral part of knowledge discovery. Learning techniques may be supervised or unsupervised. Supervised learning techniques enjoy a better success rate as defined in terms of usefulness of discovered knowledge (Maciej Majewski, 2008). **Machine learning** is the capacity of a computer to learn from experience (i.e., data) and to extract knowledge from examples. A successful learner should be able to make general conclusions about the data it is trained on. This allows it to act appropriately in new situations.

Related works

The supervised learning of morphological rules is based on the annotated data as the alignment of orthographic words with their morpheme representations. In order to apply supervised learning methods, the data should further be extended with information about inflectional classes and features (part-of-speech tags), thus making the output of the training mechanism compatible with the input for morphological analysers

There is unsupervised learning of morphological rules. The advantage of unsupervised morphological learning is that it requires only the set of orthographic words (completely “raw data”) without any stem/affix lists or grammatical annotation. However, most of the unsupervised methods put restrictions on the number of morphemes per word, on rule complexity etc. The goal of these methods is merely splitting a word into stem and suffix without the capability of assigning any grammatical information. A number of unsupervised learning techniques have been applied: genetic algorithms (Kazakov, 1997), a minimal description length approach based on spelling of words and the set of suffixes that appear with each stem (Goldsmith, 2001), and the quasi-roots algorithm (Sheremetyeva and Nirenburg, 1999).

Many researchers have been working on similar morphological learning systems for Indian languages. One of the earlier rule based morphological analyser was developed for Tamil at CIIL (Ganesan M, 1994), S.R.Kolhe and B.V.Pawar have investigated the inductive inference of grammar of subset of Marathi, Pradipta Ranjan Ray presented a computational model for Bengali Morphological analysis using Finite State methods. A generic architecture for morphological generators of agglutinative languages has been proposed (Uma Maheshwar Rao.G, 2006), in which corpus is used to extract fully inflected word forms. Stochastic taggers like HMM will not give a high accuracy for Tamil because the language is inflectionally rich and is relatively free-word order (Arulmozhi et.al 2006). In their paper, they have discussed the algorithm for developing a rule based tagger for Tamil. No work has been reported on machine learning techniques using Artificial neural network for morphology of Indian languages. ANN is applied for text reasoning (Maciej Majewski et.al, 2008), Natural language processing tasks (Joao Luis Garcia Rosa, 2002),(J.L Elman, 1990), (Ahmed, 2002),(Rajan K et.al, 2002),(Quing Ma, 2003) and categorization of Tamil documents (Rajan K. et al, 2009).

Tamil morphology

The Tamil morphology is characterised as agglutinative or concatenative. That is, morphs are agglutinated or concatenated in a sequence after a stem. Concatenative morphology in Tamil involves always suffixation (Thomas Lehman, 1993). This is represented as

Stem + (affix)ⁿ

where the superscript n means one or more occurrences of a suffix. Morphs are concatenated as suffixes at the right of a word stem, to produce inflected or derived forms of words. In Tamil, there is another morphological process which is reduplication. All lexical or root morphemes are grouped into four major types. They are verbal, nominal, adjectival and adverbial roots. All other types of words can be identified as an inflected or uninflected form of these stems. Nouns can be inflected for the case and number. An inflected noun form may be the realisation of three morphemes, as given in the following representation.

Noun stem + [Plural suffix] + [Oblique] + [Case suffix]

Verbs can be inflected for tense, person, number, gender and others.

Verb stem + [Tense Marker] +[Verbal Participle Suffix] + [Auxiliary verb] +[Tense Marker]+[Person, Number, Gender]

Postpositions and adjectives cannot be inflected. The derivation occurs only in nouns. When morphemes or words combine, certain morphophonemic changes occur. Nouns can be inflected for the case and number. Inflected verbs in Tamil are finite or non-finite. Finite verbs mark both tense and subject-verb agreement, non-finite verbs do not. Finite verbs occur only in restricted contexts in the structure of a sentence; they typically mark the end of a sentence.

The lexical knowledge is characterised with respect to verbs. All the grammatical information that is, about tense, number, gender, etc is carried by the verbs in a sentence. These are the reasons why the study of verbs has acquired immense importance in Linguistics. Modern Tamil has three tenses. Each tense morpheme is realised by a number of tense suffixes or allomorphs. Tense suffixes are listed in Table 1.

Tense	Suffixes
Present	kiR, kkiR, kinR, kkinR
Past	t, tt, ndt, in, inR
Future	p, v, pp

Table 1. Tamil tense markers

The second inflectional suffix after the stem is a PNG marker. The table 2 shows the person, number and gender suffixes.

op	Number	Gender	Suffix	
First	Singular		Een	
	Plural		Oom	
Second	Singular		Aay	
	Plural		Iirkal	
Third	Singular	Masculine	aan	Aar
		Feminine	aaL	
		Neuter	Atu	
	Plural	Masculine/ Feminine	aarkaL	
		Neuter	Ana	

Table 2. Tamil PNG markers

Tamil distinguishes between 4 types of non-finite verb forms. The non-finite verb forms, except the infinitive have both positive and negative forms. Only the adjectival participle distinguishes tense. All other non-finite verbs are tense less. Each of the non-finite verb forms is marked with a non-finite verb suffix, which is added to the verb stem, or to the tense suffix. The infinitive is formed by the affixation of infinitive suffix to the verb stem. The conditional verb form occurs both in a positive and negative forms. This is formed by adding the phoneme cluster of the past tense allomorph to the verb stem and then affixing the conditional suffix(-aal/-aaviTTaal). Adjectival participle is formed by adding past or present tense allomorphs to the verb stem and the adjectival suffix (-a).

Non-Finite	Suffixes	
Infinitive	~a	
Adj.Participle	t,tt,ndt,in,inR kiR, kkiR,um	~a
Verbal Participle	t,tt,ndt,in	~u, ~i
Conditional	~aal,~aaviTTaal	

Table 3. Non-Finite verb suffixes

The suffix ~um morph realised the future tense morpheme and the adjectival morpheme. It is an instance of homophony of morphs in Tamil. The list of morphemes is given in the Appendix A.

Neural Network

Artificial neural networks are computational models based on biological neural networks. They can be used to model complex relationship between inputs and outputs or to find pattern in data. Neural network based approaches learn the associations of word-to-tag mappings, from a training data set and also generalise to unseen examples. Neural networks have self learning capability, are fault tolerant and noise immune.

In this paper, a three layer feed forward neural network with hyperbolic tangent (tanh) function is used in hidden layers followed by a linear output layer. The neural network is trained using backpropagation algorithm. A momentum term is used to achieve a faster global convergence. A bias value is used to enable each neuron to fire hundred percent.

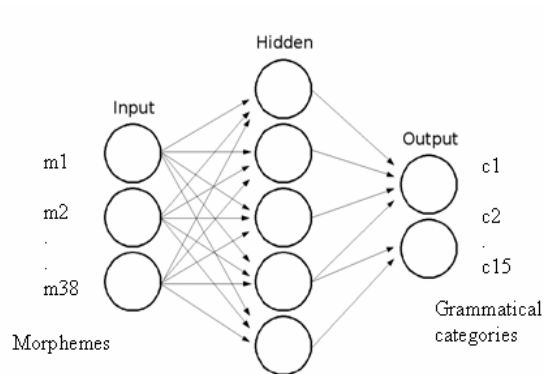


Figure 1. Architecture of ANN model

The input layer is used to represent the 38 morpheme features of the input word. The output layer represents the grammatical values of the cluster of morpheme features. The structure of the ANN model is **38 L - 30 N - 15 L**, representing the number of neurons/units in the input layer, hidden layer and the output layer respectively. The letters L and N denote linear and non-linear activation functions. In this study, the network is trained with a learning rate of 0.01 and 2000 epochs.

Feature Extraction

The input and output vectors are prepared from the word-forms generated from 208 Tamil verb stems using the morphological generator. The verbs are inflected for tense, person, number and gender. They are also inflected for non-finite forms. During the process of generating the various conjugations, their corresponding grammatical categories are also collected as output. The stems are combined with different morphemes of various categories by the generator. The morpheme table has 38 features. The categories are marked for 15 grammatical values according to the morpheme components added with the stem. This method has been utilised because, the corpus may not have all the combinations. Totally 7800 different word forms have been generated and stored in the form of input and output vectors. Among the 7800

samples, 5000 samples are randomly selected for training. The network is tested by using the remaining list of 2800 samples.

The binary representation of input and output vectors are created for all word-forms by scanning through all morphemes. The morphemes are listed in the descending order of their length. If a particular morpheme is found within the word-form, it is represented by 1 in the input vector and also the corresponding category value is set to 1 for the output vector. During testing phase, only the input vectors are created from the given word-form. The feature vector for the word is given by equation 1.

$$w = \{ m_1, m_2, m_3, \dots, m_n \} \text{ -----} \rightarrow 1$$

The word w is a sequence of n morphemes

Where $m_i = 1$ if i th morpheme is present in
the word w ., otherwise 0.

The input feature extraction from word forms during the test phase is same as the one followed during the training.

Results and Discussions

The proposed model of supervised learning for Tamil morphology shows the ability of neural networks to learn the morpheme to category association.

Number of test Samples	Numbers Correctly recognized	Precision
Total 2800	2694	96.21 %
Finite Verbs 1900	1865	98.15 %
Non-Finite Verbs 900	829	92.11%

Table 4. Performamnce of the ANN Model

The system produces correct categories for the morpheme sequences of all the inflected forms of finite verbs that are not part of the training set. The non-finite forms shows lower recognition rate because of the shorter morphemes than for the finite verb forms. The table 4 shows the performance of the model on these two verb types.

Conclusion

The proposed neural network model is a powerful supervised machine learning system that can be applied effectively for any inflectional morphological language. This approach is cost effective, because it does not require any explicit grammar rules. The experimental result shows its promising performance for morphologically rich languages like Tamil. This work can be extended to cover other grammatical categories of Tamil, by changing the input and output vectors of the neural network

References

1. Ahmed, S. Bapi Raju, P.V.S. Chandrasekhar, M. Krishna Prasad. 2002. *Application of Multilayer Perceptron Network for Tagging Parts-of-Speech*. Proceedings of the Language Engineering Conference (LEC'02), IEEE Computer Society: 57-63.
2. Arulmozhi P, Sobha L and Kumara Shanmugam B. 2004. *Parts of Speech Tagger for Tamil*. Symposium on Indian Morphology. Phonology & Language Engineering, March 19-21. IIT Kharagpur. :55-57.
3. Elman J.L. 1990 *Finding structure in Time..* Cognitive Science 14 :179-211.
4. Ganesan.M. 1994. A scheme for grammatical tagging of Corpora in Indian languages. B.B.Rajaprohit(Ed.),CIIL.
5. John A. Goldsmith. 2001. *Unsupervised learning of the morphology of a natural language*. Computational Linguistics, 27(2):153{198.)
6. Kazakov, D..2000. *Achievements and prospects of learning word morphology with inductive logic programming*. In Cussens & Dzeroski (Ed.), Learning Language in Logic. : 89-109.
7. Rumelhart D.E , J. L. McLlland. 1986. *On learning past tenses of English verbs*. In D. E.Rumelhart and J. L McLlland, editors, Parallel Distributed Processing, volume 2, pages 216{ 271. MIT Press, Cambridge, MA.
8. Rajan.K, Ramalingam.V, Ganesan M 2002a. *Corpus Analysis and Tagging*. Symposium on Translation support system. IIT, Kanpur.
9. Rajan.K, Ramalingam.V, Ganesan M 2002b, *Applications of Neural Network for Tamil Studies*. Proceedings of the International Conference on Tamil Computing. Chennai, India: University of Madras.
10. Thomas Lehman. A grammar of modern Tamil. 1993. Pondicherry Institute of Linguistics and Culture.

Syntactic Parsers for Tamil

Dr. M. Ganesan

Professor of Linguistics

Annamalai University

Introduction

Natural Language Parsing (NLP) is a process by which any natural language text is analyzed for its components in terms of sentences, clauses, phrases, and words using the grammatical rules and the instructions for operating those rules. The major objectives of the parsing are, first to identify the grammatical function of different components in a sentence and second, to understand the semantic roles of different grammatical components. Efforts are being made to achieve the goal through a number of methods. Some of them are grammatical rule based and some are statistical approach. A number of grammatical models like Phrase Structure Grammar, Transformational Generative Grammar, Generalized Phrase Structure Grammar, Head Driven Phrase Structure Grammar, Lexical Functional Grammar, Case Grammar, Minimalist Programme, etc. are being used to represent the grammatical knowledge to the system. The choice of model mainly based on the grammatical system of the language in question. As most of the time the rule-based approach does not handle some of the exceptional structures in a language, and takes more time to represent the grammatical knowledge to the system, some of the statistics-based models like probability model, Neural network model, Hidden Markov model, Support Vector model, etc. are used. Text processing, understanding, and generation are the basic needs to achieve the long term goals like Machine Translation, Man-machine communication through a natural language, Text to Speech and Speech to Text system generation, Information Retrieval, Information Extraction, Question Answering system, etc. In this paper I discuss the criteria that I evolved for identification of various phrases, clauses and sentences. The syntactic parsers that I developed work fairly well on modern Tamil corpora. The parser takes the text, which are tagged at word level, as input. The POS tags, which I have used are the standard to this parser.

Morphological Analysis

Tamil is an agglutinative language. The Morphology of Tamil is more complex. For example, a regular verb can be conjugated to around 1600 word forms. Therefore the composition of a sentence can comfortably studied at two levels: morphological and syntactic level. In morphological Analysis the main tasks for the analyzer are 1) to identify the boundary of a word, 2) to identify various morphs that constitute a word, 3) to label the grammatical category to each morph, and 4) to mark the Parts of Speech to the word. In order to achieve all these a set of Machine Readable Dictionaries (MRD) for stems and suffixes, and the morphotactics (i.e. the arrangement of morphs in different word forms) are basically needed. They represent the morphological information to the system for analysis words (For more details see Ganesan, 1994). The output of the Morphological Analyzer will be a word-level tagged text.

The morph level tags are marked in between two underscores immediately after the morphs. And the word level tags are marked inside an angular bracket '< >' immediately at the end of the word.

For example

avaL_pro_<PNh> paaTTu_abn_<NN> paaTinaaL_vb, pst, 3sf_<NN>

'She sang a song'

Syntactic Analysis

Normally, the syntactic parser makes use of grammatical rules and a dictionary consisting of words and their grammatical information. In any parsing programme, the lexical items within a sentence are first marked with their grammatical value, making use of the dictionary. Then, using different grammatical models and grammatical rules, the system parses a sentence and marks the phrases and clauses in the sentence.

The tasks involved for a syntactic parser are:

- (i) to identify a sentence in a text
- (ii) to identify various phrases and marking them in a sentence
- (iii) to identify the head of a clause
- (iv) to segment a clause in a sentence and
- (v) to mark a clause with the grammatical information

For the identification of both clauses and sentences, first the text have to be tagged at phrase level. The phrase level tagging includes all the phrases, such as, noun phrase, verb phrase, adjectival phrase, adverbial phrase, quantifier phrase and postpositional phrase. The criteria for identifying the phrases are explained in detail in the report (Ganesan, 2010). For example, there are 14 rules given for the identification of different noun phrases. With these rules the system identified 98% of noun phrase correctly in the given text. The system fails only when two <NN>s marked for Nominative case occur consecutively. To solve such problems the system needs extra linguistics knowledge, which is difficult to provide in the knowledge base. Statistics-based approach can be used to resolve such problem. In this paper the criteria for identification of clauses and sentences are only discussed. We mark broadly five types of clauses in a sentence. They are:

- i) Se - Embedded sentence
- ii) Sm - Matrix / Main sentence
- iii) Si - Interrogative sentence
- iv) EQS - Equational sentence
- v) Sel - Elliptical sentence

Basic Principles

The parsing described here is built on the principles explained below for the analysis of a sentence. They are:

- i) The head of a clause always finds its place in the right most position of a clause. In other words, clauses in Tamil are left-branching.
- ii) In a complex sentence where we have more than one clause, the arguments for a predicate (verb) occur before the verb of that clause and never in another clause.

Approach

For syntactic analysis, a sentence is the minimum unit. Identification and marking of a sentence is not very simple. Most of the time, the sentences are identified using the punctuation markers like full stop (.), question mark (?), colon (:), semicolon (;), etc. apart from these punctuation marks, in a quotation, the quoted sentence and the quoting constitute are two different sentences. In the present study, a set of algorithm is used in the parser to identify a sentence.

Once a sentence is identified, the parser looks at the sentence word by word. This can be done either from the starting of a sentence or from the end of a sentence. In other words, a sentence can be approached either from Left-to-Right (L to R) or Right-to-Left (R to L). In the morphological analysis the words are approached from R to L. But in Syntactic parsing the sentences are approached from L to R. The reasons are as follows. The difference between the formation of a word and a sentence is that in word the suffixes are added to the right of the stem, whereas in sentence the arguments are added to the left of the head of a phrase or clause. The suffixes to a stem and the arguments to a head of a phrase / clause are dependent, but the stem (root/base) of a word and the head of a phrase / clause are independent unit. The dependent unit indicates the parser to proceed till it gets a correct independent unit. Therefore it is always easy and convenient to start with a dependent unit.

Procedures

As the finite State Automata (FSA) and the Augmented Transition Network (ATN) advocate, the input strings are taken for analysis, one by one, keeping the starting node as the first word of a sentence. The parser proceeds further, making use of the grammatical information labeled to the words and traverses to the next state. In our approach based on the word level grammatical tags attached with the input text, the parser select different modules. For example, if the word has <NV> as its tag, the parser moves to non-finite verb modules and identifies it as a clause 'Se'. In each module, the parser follows two types of procedures.

- i) First, it looks for the information within the word to decide a clause.
- ii) If the information is not sufficient or if the information indicates to look at the following word, then the parser takes the information from the next word.

Identification of clauses and Sentences

All the four non-finite form of verbs viz., infinitive, conditional, verbal participle and relative participle constitute a sub-ordinate clause. In addition to these affirmative constructions, the negative forms (except infinitive, which does not have negative form) also construct the sub-ordinate clauses. Since the relative participle has the adjectival function, it is not segmented as a clause. As mentioned in the previous section of this paper, there are 5 types of sentences, including the subordinate clause.

Embedded Sentence

In the present study, the subordinate clause is marked as 'Se' that is, "sentence embedded". The criterion for identifying the sentence embedded is given below:

- i) **If a verb phrase (VP) is marked as conditional (VP-con), or verbal participle (VP-vp) or infinitive (VP-inf), then the text from the beginning of the sentence or after a <NV> verb, till the VP is as a sub-ordinate clause (sentence embedded)**

For example

1. *raaNi nanRaakap paTittu veelaikkuc cenRaal*

[raaNi < NN: (ppn)>] NP_nom [nanRaaka < AV: (adv) > [paTittu < NV: (vpm, pst)>]] VP_vp
[veelaikku < NN: (ian, dat)>] NP_dat [cenRaal < FV: vb, pst, 3sf>] VP.

'Rani studied well and went to a job'

Since a non-finite form, verbal participle (VP_vp) has occurred with verb phrase

[[raaNi < NN: (ppn)>] NP_nom [nanRaaka < AV: (adv)> [paTittu < NV: (vpm, pst)> VP_vp]] Se 'Rani having studied well'

is the sub-ordinate clause.

Main Sentence (Verbal)

The main clause of a sentence is called main / matrix sentence and is marked as 'Sm'. The criterion for identifying the main clause is:

- ii) **If a verb phrase (VP) is not marked for any of the non-finite verb as _inf_, _con_, _vp_, then mark it as main clause.**

For example

2. *mallikaa neeRRu viiTTukku poonaaL*

[mallikaa < NN: (ppn)>] NP_nom [viiTTukku < NN: (ian, dat)>] NP_dat
[neeRRu < AV: (adv)> poonaaL < FV: Vb, pst, 3sf>] VP

'Malliga went home yesterday'

Since, the verb phrase is not marked for _inf_, or _con_, or _vp_, [[mallikaa < NN: (ppn)>] NP_nom [viiTTukku < NN: (ian, dat)>] NP_dat [neeRRu < AV: (adv)>] poonaaL < FV: Vb, pst, 3sf>] vp] Sm, 'Malliga went home yesterday' is the main sentence.

Interrogative Sentence

To identify the interrogative sentence, the following mechanism is adopted.

- iii) **If an interrogative word <ITW> or an interrogative marker occurs in a sentence, treat that sentence as an interrogative sentence.**

For example

3. *raaman een azutaan*

[raaman < NN: (ppn)>] NP_nom [een < ITW:] [azutaan < FV: (vb, pst, 3sm)>] VP.

'Why did Rama cry?'

Since, an interrogative word <ITW> has occurs, [[raaman < NN: (ppn)>] NP_nom [een < ITW: (itw)>] [azutaan < FV: (vb, pst, 3sm)>] VP] Si 'Why did Rama cry?' is an interrogative sentence and is marked as 'Si'.

4. *ciitaa koovilukkuc cenRaaLaa?*
 [*ciitaa* <NN: (ppn) >]NP_nom [*koovilukku* <NN: (ian, dat) >] NP_dat
 [*cenRaaLaa?* <FV: (vb, pst, 3sf, int)>] VP.

'Did Sita go to temple?

Here, an interrogative marker -aa has occurred with the finite verb, hence the sentence is marked for 'Si' and is given below:

- [*ciitaa* <NN: (ppn) >]NP_nom [*koovilukku* <NN: (ian, dat) >] NP_dat
 [*cenRaaLaa?* <FV: (vb, pst, 3sm, int)>]VP]Si.

Equational Sentence

The criterion for identifying the equational sentence (EQS) is given below:

- (vi) If a sentence has only two noun phrases (NP) and are marked for nominative case (NP_nom), mark the sentence as an equational sentence.**

For example

5. *atu cari*
 [*atu* <PNn: (pro, dem) >] NP_nom [*cari* < NN: abn>] NP_nom

'That is right'

Elliptical Sentence

The mechanism adopted to identify the elliptical sentence 'Sel' is

- (v) If one or more than one noun phrase (NP) is marked for any case or for _emp_, mark it as an elliptical sentence.**

For example

6. *aamaam cengkatir*
 [[*aamaam* < NN: (ind)> cengkatir<NN: (ppn)] NP_nom] Sel
 'Yes Senkathir'

The above sentence has only one noun phrase marked for nominative case and hence it is an elliptical sentence and is marked as 'Sel'.

Conclusion

As the parser has already identified the phrases and they are marked properly, it is easy to mark the clause and sentence boundary. With the above listed rules the system is capable of identifying 100% of clauses and 97% of the sentences in the given sample. The study was carried out on a smaller corpus. When a large corpus is taken for analysis, few more rules may be needed to be incorporated in the system.

Reference

1. Andropov, M. 1965. **The Tamil language**. Moscow: 'Nauka' Publishing house.
2. Ganesan, M. 1994 "A Scheme for Grammatical Tagging of large Corpora" in B.B. Rajaprohit (Ed.) **Technology and Languages** Mysore: Central Institute of Indian Languages.
3. Ganesan, M. 2003. "Computational Grammar of Tamil". In B.Rama Krishna Reddy (Ed.) **Word Structure in Dravidian**. Kuppam: Dravidian University.
4. Ganesan, M. 2005. **Corpus** Analysis Tools for Tamil (CATT) Ver: 1.0. (Tamil Software) Annamalai Nagar: Wisdomsoft Publications.
5. Ganesan, M. 2010, UGC Major Research Project Report on " Syntactic Parsers for Tamil" Annamalai nagar: Annamalai University.
6. James, Allan. 1995. **Natural Language Understanding**. United states of America: The Benjamin / Cummings Publishing Company.
7. Jurafsky, Daniel. and H. James Martin. 2000. **Speech and language Processing**. Singapore: pearson Education.
8. King, M. 1983. **Parsing Natural Languages**. London: Academic Press.

கணினியில் தமிழும் பிற திராவிட மொழிகளும்

முனைவர் **இராதா செல்லப்பன்**

சிறப்புநிலைப் பேராசிரியர்

பாரதிதாசன் பல்கலைக்கழகம்

திருச்சிராப்பள்ளி - 620 024.

மின்னஞ்சல் : radcell@yahoo.com

தமிழ், தெலுங்கு, கன்னடம், மலையாளம் முதலான திராவிட மொழிகளில் கணினிப் பயன்பாடு பல ஆண்டுகளாக நடைபெற்று வருகின்றன. கணினியில் மட்டுமின்றி இணையத்திலும் திராவிட மொழிகள் உலா வருகின்றன. தமிழில் நடைபெறும் கணினி, தகவல் தொழில்நுட்பச் செயல்பாடுகள், சிக்கல்கள் தொழில்நுட்ப வளர்ச்சி போன்றவை இணையத் தமிழ் மாநாட்டில் விவாதிக்கப்பட்டு வருகின்றன. ஆனால் பிற திராவிட மொழிகளில் இது போன்ற மாநாடுகள் அதிகமாக நடைபெறுவதாகத் தெரிவித்தல்லை. எனவே பிற மொழிகளில் நடைபெறும் இது போன்ற செய்திகளை நாம் அறியும் வாய்ப்பு குறைவு. எனவே பிற திராவிட மொழிகளில் நடக்கும் கணினி முயற்சிகளை அறிவது நன்மை பயக்கும் என்ற ஆர்வத்தில் இக்கட்டுரை ஆக்கப்படுகிறது. திராவிட மொழிகள் என்றால் பெரும்பான்மையோரால் பேசப்பட்டு வரும் தமிழ், தெலுங்கு, கன்னடம், மலையாளம் ஆகிய நான்கினைப் பற்றிய தகவல் தொகுப்பாக இக்கட்டுரை அமைக்கப்பெற்றுள்ளது.

கணினித் தொழில்நுட்ப மொழி வள மையங்களும் அவற்றின் மென்மப் பணிகளும்

2000-ஆம் ஆண்டு இந்தியாவின் தகவல் தொழில்நுட்ப அமைச்சகத்தின் ஒரு பிரிவான இந்திய மொழிகளில் தொழில்நுட்ப மேம்பாட்டுப் பிரிவு (The Technology Development for Indian Languages (TDIL) 13 மையங்களை பல கல்வி நிறுவனங்களிலும் அமைத்தது. இவையனைத்தும் இந்திய மொழித் தீர்வுக்கான தொழில்நுட்ப வள மையம் (Resource Centre for Indian Language Technology Solutions-RCILTS) என்ற பெயரால் அறியப்பட்டன. தேவைக்கேற்ற முறையில் புதிய மென்மங்களை உருவாக்குதலும் அவற்றை வெளியிடுதலும் என இருநிலையில் அவை செயலாற்றி வருகின்றன. இந்த வள மையங்கள் ஒவ்வொன்றுக்கும் இப்பணிக்காக ரூபாய் ஒரு கோடி அளிக்கப்பட்டன. அந்தந்த மொழி மாநிலத்தில் இவ்வள மையங்கள் அமைக்கப்பட்டன. தமிழுக்காக அமைக்கப்பட்ட மையம் **அண்ணா பல்கலைக்கழகத்தில்** இயங்குகிறது. இந்த ஆய்வு மையம் அண்ணா பல்கலைக்கழக **கே.பி.சந்திரசேகர் ஆய்வு மையம்** என அழைக்கப்படுகிறது. தெலுங்கு மொழிக்கான வள மையம் **ஹைதராபாத் பல்கலைக்கழகத்தில்** அமைந்துள்ளது. கன்னட மொழிக்கான வள மையம் **பெங்களூர் இந்திய அறிவியல் கழகத்தில்** அமைந்துள்ளது. இது எல்லா கன்னட அபிமானிகள் அந்தர்ராஷ்டிரிய வேதிக இன்டர்நேஷனல் E-Kavi (Ella Kannadaabhimaniigala Antharrashtriya Vedike International). என்பதாகும். மலையாள மொழிக்கான மையம் COWMAC (Consortium of Organizations Working on Malayalam Computing).

எழுதி, உரை பதிப்பி, சொற்செயலி, அகராதி, மின்னகராதி, அகராதி மென்மம், சொல் வலை, சொற்றொடர் அடைவு, உருபன் பகுப்பாய்வி (Morphological Analyzer), தரவுக் களம் உருவாக்குதல் (Corpus Development), எந்திர மொழிபெயர்ப்பு (Machine Translation), எழுத்துணரி (Text to Speech-TTS), பேச்சுணரி (Speech to Text-STT), தொகுப்பி (Summarizer), தகவல் மீட்டல் (Information Retrieval), ஒளியெழுத்துணரி (OCR) தேடிகள் (Search Engine in Regional Languages), கணினி வழிக் கற்றலும்

கற்பித்தலும் Computer Aided Language Learning முதலான பல நிலைகளில் கணினி மென்மங்கள் பலவும் திராவிட மொழிகளில் உருவாக்கப்பட்டுள்ளன.

எழுதிகள் (Script Processors)

தமிழ் எழுத்துருக்களை உருவாக்குவதில் நிறுவனங்களுக்கு மட்டுமன்றி தனியார்களுக்கும் பெரும் பங்குண்டு. பேரா.கல்யாண சுந்தரத்தின் மயிலை எழுத்துரு, திரு சீனிவாசனின் ஆதமி, திரு இரவீந்திரனின் துணைவன், திரு கோபாலகிருஷ்ணனின் புதுச்சேரி எழுத்துரு, திரு ஜார்ஜ் ஹார்ட்டின் காஞ்சி முதலானவை தொடக்ககால முயற்சிகள். அமுதம், சுரபி, பாமினி முதலானவை தமிழ்நாடு மட்டுமன்றி உலகம் முழுவதும் பரக்கப் பயன்பட்ட எழுத்துருக்கள். TAM, and TAB ஆகிய குறியீட்டு முறைகள் தமிழக அரசால் அங்கீகரிக்கப்பட்டவை. மாடுலர் சிஸ்டம்ஸ் அங்குர் மென்மத்தைப் பதினொரு இந்திய மொழிகளுக்காக உருவாக்கியது. இன்ஸ்கிரிப்ட் என்பது இந்திய மொழிகளுக்காக சாஃப்ட் வியூ நிறுவனத்தால் உருவாக்கப்பட்டது. இது போன்ற மற்றொரு மென்மம் சித்திரலேகா. கம்பன், முரசு போன்ற பல மென்மங்களும் தமிழ்க் கணினியாளர்களிடையே நல்ல வரவேற்பினைப் பெற்றன. யுனிகோடு எழுத்து முறையில் பல்வேறு தட்டச்சு முறைகளிலே யுனிகோடு குறியீட்டினைத் தட்டச்சு செய்யும் முறையில் சாதனைபுரிந்து காட்டிய என்.எச்.எம் (NHM) நிறுவனத்தின் பணி குறிப்பிடுதற்குரியது. ஜிஸ்ட்(GIST) தொழில்நுட்பத்தில் சிடாக் நிறுவனம் ஒரு சொற்செயலியை உருவாக்கியது. இந்திய மொழிகளுக்கான ஒரு மென்மமும் சிடாக் -ஆல் உருவாக்கப்பட்ட ஸ்ரீலிபியாகும். பல எழுதிகளும் தொடக்க காலகட்டத்தில் பிற திராவிட மொழிகளிலும் உருவாக்கப்பட்டுள்ளன.

சொற்செயலிகள் (Word Processors)

சிடாக் நிறுவனத்தால் உருவாக்கப்பட்ட ஐ-லீப் தமிழ், தெலுங்கு, கன்னடம், மலையாளம் உட்படப் பல இந்திய மொழிகளுக்குமாக உருவாக்கப்பட்ட உரை பதிப்பியாகும். கணினிப் பயன்பாட்டிற்கு மட்டுமன்றி வலைப் பயன்பாட்டிற்கும் இந்த மென்மம் பயன்பட்டது.

இந்திய மொழிகளின் நடுவண் நிறுவனம் பாரதி என்கிற சொற்செயலியை வெளியிட்டது. மற்றொரு செயலியான இளங்கோ தவறான சொற்களை அடிக்கோடிட்டுக் காட்டும். அகராதியில் புதிய சொற்களைச் சேர்க்கும். சந்திப் பிழைகளைக் கண்டறியும். சக்தி ஆங்கிலம்-தமிழ் ஆபிஸ் மென்மம் நல்லதொரு சொற்செயலியை உள்ளடக்கியது. குரல் செயலி என்பது ஒரு சொற்செயலி. இதில் தமிழ் உரையை ஒலியாக மாற்றித் தரும் வசதியும் உண்டு. அண்ணா பல்கலைக்கழக வள மையம் (RCILTS), பலகை என்ற இருமொழிச் சொற்செயலியை உருவாக்கியது undo and redo செயல்பாடுகளைக் கட்டுப்பாடின்றி பயன்படுத்தலாம். பொதுவாக. இச்சொற்செயலிகளில் எச்.டி.எம்.எல் கோப்புகளைப் பார்வையிடும் வசதியும் மின்னஞ்சல் அனுப்பும் வசதியும் இருந்தன. இந்த மென்மத்தில் சொல்லெழுத்துத் திருத்தியும் இலக்கணத் திருத்தியும் இருந்தன. இன்சாஃப்ட் நிறுவனம் தமிழுக்கான ஒரு பிழை திருத்தியை உருவாக்கியுள்ளது. இதுபோன்றே பிற நிறுவனங்களும் தனியார் முயற்சிகளும் பிழை திருத்தியை உருவாக்கியுள்ளன.

மலையாள வள மையம் ISCII அடிப்படையில் நேர்ப்பதம் என்ற மென்மத்தை உருவாக்கியது. இம்மென்மம் சொல்லெழுத்துக்களில் உள்ள பிழைகளைச் சுட்டுவதோடல்லாமல் பரிந்துரைச் சொற்களையும் தந்தது. மலையாளத்திலுள்ள மற்றொரு மென்மம் அக்ஷரமாலா. இதில் இரண்டு கட்டுகள் உள்ளன. ஒன்று மலையாளம் எழுதி மேலாண்மை MSM (Malayalam Script Manager), மற்றொன்று வாகினி. இது மலையாள எழுத்துருவிற்கான கட்டு. இந்தப் பயனிலை மென்மங்கள் விசையச்சு மேலாளர் ஒன்றும் கொண்டது. இது INSCRIPT விசைப் பலகையும் கொண்டது. வாகினி கட்டு என்பது மலையாள மொழி எழுத்துருக்கள், ஆங்கிலம்-மலையாளம் எழுத்துருக்கள், வலை எழுத்துருக்கள் முதலியனவற்றைக்

கொண்டது. மலையாளத்தில் மின்னாளுகை நோக்கில் **நிலா** என்ற மென்மம் தயாரிக்கப்பட்டுள்ளது. இதனைக் கிளிக் (Centre for Linguistic Computing keralam) என்ற நிறுவனம் தயாரித்துள்ளது.

தெலுங்கு வள மையம் **அஷ்டரா** என்கிற பன்மொழிச் சொற்செயலியை உருவாக்கியது. இதனை எல்லா தளங்களிலும் பயன்படுத்த இயலும்.

சூப்பர் சாஃப்ட் என்பது (Supersoft) கேரளத்தில் உள்ள ஒரு கணினி மென்ம மையம். இம் மையம் 1990-களிலிருந்தே இந்திய மொழிகளுக்கான மென்மத்தை உருவாக்குவதில் ஈடுபட்டுள்ளது. இந்தியாவின் கடைகோடியிலுள்ள மனிதர்களும் கூட கணினியைப் பயன்படுத்த வேண்டும் என்ற இந்திய நோக்கில் இம் மையம் செயல்பட்டது. 1991-இல் முதல் மலையாளம் சொல் செயலி இம் மையத்தால் வெளியிடப்பட்டது. 1994-இல் **தூலிகா**(Thoolika) என்ற மென்மமும் இம்மையத்தால் வெளியிடப்பட்டது. இது வின்டோஸில் இயங்கியது. தூலிகாவின் பல பதிப்புகள் அடுத்தடுத்து வெளிவந்தன. தூலிகா 2005-இல் மலையாளம் இந்தி தமிழ் உரைகளைத் தட்டச்சு செய்யும் வசதியும் ஏற்படுத்தப்பட்டது. தூலிகா 2006 தகுதர எழுத்துருவையும் உடையதாக வடிவமைக்கப்பட்டது. இம்மென்மம் பழைய மலையாள எழுத்துக்களையும் எழுத்துச் சீர்திருத்த எழுத்துக்களையும் உடையதாக இருந்தது. பழைய எழுத்துகளிலிருந்து புதிய எழுத்துக்களுக்கு மாறும் வசதியும் செய்து தரப்பட்டுள்ளது. இதில் கேரள அரசால் அங்கீகரிக்கப்பட்ட இன்ஸ்கிரிப்ட் விசைப்பலகை தட்டச்சு விசைப்பலகை உள்ளன. ஒரே தட்டில் சில்லெழுத்துக்களை உள்ளிடும் வசதி செய்யப்பட்டுள்ளது. அகர வரிசைப்படுத்தும் வசதி 99 விழுக்காடு செய்யப்பட்டுள்ளது என்று அதனை உருவாக்கியவர்கள் கூறுகின்றனர்.

குளோபல் எழுதி Global Writer என்பது ஒரு சொற்செயலி. இது 100 மொழிகளை ஒரே விசைப்பலகை செய்யும் வசதியைக் கொண்டது. PDSTEXT என்பது தகுதர வலைய மென்மம். இதில் தமிழ், தெலுங்கு, மலையாளம், கன்னடம், இந்தி ஆகிய மொழிகளும் ஆங்கில மொழியும் பயன்படுத்தும் வசதி உண்டு.

எழுத்துரு மாற்றி Script Converter

தமிழில் பல வகையான எழுத்துருக்களும் உள்ளன. ஒரு குறியீட்டிலுள்ள எழுத்துருக்களைப் பிறிதொரு குறியீட்டு எழுத்துருக்களாக மாற்றும் வகையில் பல எழுத்துரு மாற்றிகள் தமிழில் உருவாக்கப்பட்டுள்ளன. அவற்றுள் குறிப்பிடத் தக்கன **பொங்கு தமிழ்** இணையவழி மாற்றி, **என்..எச்.எம்** மாற்றி, **மாறன்** மாற்றி அவற்றுள் சில. ஐஎஸ்எம் இம்போர்ட் (ISM Import) என்பது சிடாக் -ஆல் உருவாக்கப்பட்ட மாற்றி. இது பிற எழுத்துருக்களை ISFOC வடிவத்திற்கு மாற்றும்.

இணைய மின்னகராதிகள் Electronic Online Dictionaries

தென்னாசிய மின்னகராதிகள் (Digital Dictionaries of South Asia) என்ற இணையதளத்தில் பல மொழிகளின் அகராதிகள் இடப்பட்டுள்ளன. 24 ஆசிய மொழிகளின் அகராதியையும் ஆங்கில அகராதியையும் தருகிறது. அத்துடன் ஒப்பீட்டு அகராதிகளான **டி.இ.டி** யும் (Dravidian Etymological Dictionary) திராவிட மொழிகள், இந்தோ ஆரிய மொழிகளுக்கான **டர்னர் அகராதியும்** இடம்பெறுகின்றன. Charles Phillips Brown Dictionary of Telugu - English Dictionary இத்தளத்தில் காணப்படுகிறது. தமிழில் எழு அகராதிகளும் கன்னடத்திற்கு ஒரு அகராதியும் மலையாளத்திற்கு இரண்டு அகராதிகளும் தெலுங்கிற்கு நான்கு அகராதிகளும் இடம்பெற்றுள்ளன. சென்னைப் பல்கலைக்கழகத்தின் பேரகராதி இதைத் தவிர மேலும் இரண்டு வலைகளிலும் காணப்படுகின்றன. **கொலோன்** தளம். இணையப் பல்கலைக்கழகத் தளம் ஆகியவற்றிலும் இடம்பெறுகிறது. **ஆங்கிலம் - சிங்கள அகராதி**, **உமர் தமிழ் அகராதி**, **வெப் உலகம் அகராதி** முதலியன பிற தமிழ் அகராதிகள். தமிழ் வள மையம் **சொல்லோவியம்** என்றதொரு பட அகராதியைத் தயாரித்துள்ளதாக அறிகிறோம். **சொற்புதையல்** அவர்களது இணைய அகராதி. SMART (Small Readable Tamil) என்பது இந்திய மொழிகளின் நடுவண் நிறுவனத்தால் உருவாக்கப்பட்ட இணைய அகராதி. **புலவன்** (தமிழ் மொழி அகராதி), **பாலம்** (ஆங்கிலம் - தமிழ் இருமொழி) ஆகிய இரு அகராதிகளும்

பனசியா டிரீம் வீவர்ஸ் நிறுவனத்தின் தயாரிப்புகள். **பால்ஸ் மின் அகராதியும்** இங்கு குறிப்பிடத்தக்கது. பாரதிதாசன் பல்கலைக்கழகம் 1997-இல் அகராதி மென்மம் ஒன்றைத் தயாரித்துள்ளது.

கன்னட வள மையம் கன்னட மொழிக்கான சொல் வலையொன்றினைத் தயாரித்துள்ளதாக அறிகிறோம்.

தமிழ்ச் சொற்றொடரடைவுக்கான ஒரு இணையதளம் உள்ளது. ஆனால் அதில் ரோமன் வரிவடிவ ஆவணத்தை மட்டுமே பயன்படுத்த முடியும்.

உருபன் பகுப்பாய்வு

வினைச் சொற்களின் உருபனியல் பகுப்பாய்விற்கான ஒரு மென்மம் திரு வாசு ரங்கநாதனால் உருவாக்கப்பட்டுள்ளது. திரு துரைபாண்டியவர்களும் இத்தகைய முயற்சியில் ஒரு மென்மம் உருவாக்கியிருப்பதாகத் தெரிகிறது. சென்னைப் பல்கலைக்கழகத்தின் தமிழ் மொழித்துறை உருபன் பகுப்பாய்வு மென்மம் ஒன்றினைத் தயாரித்துள்ளது. அண்ணாமலைப் பல்கலைக்கழகத்தின் பேராசிரியர் முனைவர் கணேசன் தமிழுக்கான உருபன் பகுப்பாய்வு மென்மத்தைத் தயாரித்துள்ளார். **உருபன் பகுப்பி** ஒன்று தெலுங்கு வள மையத்தால் உருவாக்கப்பட்டுள்ளது. இப்பகுப்பியில் 64000 வேர்ச்சொற்கள் பயன்படுத்தப்பட்டுள்ளதாக அதன் அலுவலக வெளியீட்டுக் குறிப்பு கூறுகிறது. கன்னட மொழிக்கும் இது போன்றதொரு உருபன் பகுப்பி தயாரிக்கப்பட்டுள்ளது.

தரவுக் களம் தயாரித்தல்

இந்திய மொழிகளின் நடுவண் நிறுவனம் திராவிட மொழிகளுக்கான தரவுத் தளங்களைத் தயாரித்துள்ளது. சென்னை வள மையம் இந்தத் தரவுகளை திருத்தியமைத்தது. தெலுங்கு வள மையம் 225 நூல்களிலுள்ள ஏறத்தாழ 30,000 பக்கங்களில் 9.25 மில்லியன் சொற்களைச் சேமித்துள்ளதாகத் தெரிவிக்கிறது.

எந்திர மொழிபெயர்ப்பு Machine Translation and Translation Tools

தமிழ்ப் பல்கலைக்கழகம் 1980-களிலேயே எந்திர மொழிபெயர்ப்பு முயற்சியில் ஈடுபட்டது. ரஷிய மொழி அறிவியல் நூல்களைத் தமிழில் மொழிபெயர்க்கும் முயற்சி செய்யப்பட்டது. திரு. துரைபாண்டி ஆங்கிலம்-தமிழ் எந்திர மொழிபெயர்ப்பில் ஈடுபட்டு ஒரு மென்மத்தை உருவாக்கினார். வாசு ரங்கநாதன் ஆங்கிலம்-தமிழ் எந்திர மொழிபெயர்ப்பினை இணையத்தில் உருவாக்கியுள்ளார். எளிய வாக்கியங்களைப் பெயர்க்கும் மென்மம் இது.

கன்னடச் சொற்களை இந்தியில் மொழிபெயர்க்கும் மென்மம் **அனுசர்க்கா** கான்பூர் இந்திய தொழில்நுட்பக் கழகத்தால் உருவாக்கப்பட்டது. பின்னர் தெலுங்கு-இந்தி மென்மமும் உருவாக்கப்பட்டது. இந்நிறுவனம் **ஆங்கிலபாரதி** என்ற மென்மத்தை உருவாக்கியது. ஆங்கிலத்திலிருந்து பிற இந்திய மொழிகளுக்குச் சில குறிப்பிட்ட புலங்களில் மொழிபெயர்ப்பு செய்ய இந்த மென்மம் உருவாக்கப்பட்டது.

ஆங்கிலம்-கன்னடம், ஆங்கிலம்- தமிழ், கன்னடம்-தமிழ் மொழிபெயர்ப்பு மென்மம் **சரஸ்வதி** என்ற பெயரில் உருவாக்கப்பட்டுள்ளது.

அரசு ஆவணங்களை ஆங்கிலத்திலிருந்து கன்னடத்தில் மொழிபெயர்க்கும் நோக்கில் கன்னட அரசு எந்திர மொழிபெயர்ப்பு மென்மத்தை உருவாக்கியுள்ளது.

எழுத்துணரி Text to Speech system

உரை ஆவணத்தைப் பேச்சொலியாக மாற்றும் எழுத்துணரிகளை உருவாக்கும் முயற்சி திராவிட மொழிகளில் ஆங்காங்கே செய்யப்பட்டு வருகிறது. ஆங்கிலத்திலே **ஈ ஸ்பீக்** என்ற மென்மம் தமிழ் ஆவணங்களைப் படிக்கும் வகையில் NVDA நிறுவனத்தால் உருவாக்கப்பட்டுள்ளது. ஆனால் அந்தத் தமிழ் உச்சரிப்பு ஆங்கில உச்சரிப்புவோல அமைந்துள்ளது. தமிழ் வள மையம் **எதிரொலி** என்ற பெயரில் இத்தகைய ஒரு மென்மத்தைத் தயாரிப்பதாக அறிகிறோம். அது போன்றே பெங்களூருவிலுள்ள இந்திய

அறிவியல் கழகமும் **திருக்குரல்-1, திருக்குரல்-11** என இரு பதிப்புகளை வெளியிட்டுள்ளன. பெண்குரல், ஆண்குரல், குழந்தைக் குரல் ஆகிய மூன்று நிலைகளில் படிக்கும்படியான வசதிகள் செய்யப்பட்டுள்ளன. உரையிலுள்ள எண்களையும் அது படிக்கும். துக்கம், மகிழ்ச்சி, கோபம் முதலான உணர்வுகளையும் தொகுத்துக் காட்டுவண்ணம் இதனை வடிவமைக்கும் முயற்சியும் மேற்கொள்ளப்பட்டுள்ளது.

மலையாளத்தில் **சுபாஷினி** என்றொரு எழுத்துணரி தயாரிக்கப்பட்டுள்ளது. இது மலையாள ஆவணங்களைத் தெளிவான நிலையில் படித்துக்காட்டுகிறது. இது மலையாள வள மையத்தால் தயாரிக்கப்பட்டது மற்றொரு மலையாள எழுத்துணரி திருவனந்தபுரம் அனைத்துலக திராவிட மொழிப் பள்ளியால் உருவாக்கப்பட்டுள்ளது. இதன் முதன்மை ஆய்வாளர்களாக பேரா. வ.அய்ய.சுப்பிரமணியனாரும் பேரா. வி.பி.சௌந்தரியும் இருந்தனர். இந்தியப் புள்ளியியல் நிறுவனத்துடன் இணைந்து இம் மென்மம் உருவாக்கப்பட்டுள்ளது. 1999-இல் உருவாக்கப்பட்ட இந்த மென்மம் இரு கூறுகளைக் கொண்டது. ஒன்று தட்டச்சு செய்யச் செய்யப் படிக்கும் கூறு. மற்றொன்று பிரெய்லி தட்டச்சிற்கேற்ற கூறு. பிரெய்லி முறைக் கூறில் எல்லா விசைகளும் பயன்படுத்தப்படுவதில்லை. நடுவிலுள்ள ஆறு விசைகள் மட்டுமே தட்டச்சு செய்யப் பயன்படுகின்றன. இதன்மூலம் பார்வையற்றோர் விசைப்பலகையில் தட்டச்சு செய்யும்போதே அந்த உரையைக் கேட்கலாம். இந்நிறுவனம் தமிழுக்கும் இத்தகையதொரு மென்மத்தை உருவாக்கியுள்ளது. ஆனால் தமிழ் மென்மத்தில் பிரெய்லி முறைத் தட்டச்சு சேர்க்கப்படவில்லை.

பேச்சுணரி Speech to Text

பேச்சொலியை உரையாக மாற்றித் தரும் முயற்சியும் திராவிட மொழிகளிலே செய்யப்பட்டு வருகின்றன. திராவிட மொழிகளில் இத்தகைய மென்மம் எதுவும் உருவாக்கப்பட்டதாகத் தெரியவில்லை. ஆங்கிலத்திற்குப் பல மென்மங்கள் உருவாக்கப்பட்டுள்ளன.

தகவல் மீட்டல் Information Retrieval

அன்வேஷக் என்பது இயற்கை மொழி உரைகளிலிருந்து தகவல்களை மீட்டெடுக்கும் ஒரு மென்மம். இது சிடாக் நிறுவனத்தால் தயாரிக்கப்பட்டது. குறிப்பிட்ட தகவல்களைத் தானாகவே உரைகளிலிருந்து எடுத்துக் கொடுக்கும்வண்ணம் அமைக்கப்பட்டது இம்மென்மம். இந்த மென்மத்தைப் பயன்படுத்தி குறிப்பிட்ட தகவலை ஆவணத்திலிருந்து பிரித்தெடுக்க இயலும். "A Knowledge Engineering Approach is being used for the development of Anveshak. Grammars are constructed by hand, domain patterns are discovered by a human expert through introspection and inspection of a corpus. Much laborious tuning and "hill climbing" statistical methods are used" என்று இதனை உருவாக்கிய சிடாக் நிறுவனம் குறிப்பிடுகிறது.

ஒளியெழுத்துணரி OCR

ஒளியெழுத்துணரி OCR (Optical Character Recognition) என்பது வருடப்பட்ட ஆவணங்களை உரை உதவும் ஒரு மென்மம். சிடாக் நிறுவனம் OCR மென்மம் **சித்ராங்கள்** என்ற ஒரு இந்திய மொழிகளுக்கான ஒளியெழுத்துணரியை உருவாக்கியுள்ளது. தமிழ் மொழிக்கான ஒளியெழுத்துணரி மென்மத்தை உருவாக்குவதில் முன்னோடியாகத் திகழ்பவர் பேரா.கிருஷ்ணமூர்த்தி ஆவார். அவரது ஒளியெழுத்துணரிக்குப் பொன்விழி என்று பெயர். இக்னூ நிறுவனம் அச்சு நூல்களுக்கான ஒளியெழுத்துணரியைத் தயாரித்துள்ளது. மலையாளம் வள மையம் **நயனா** என்ற ஒளியெழுத்துணரியை உருவாக்கியுள்ளது. மலையாளத்திலுள்ள வருடப்பட்ட உரைகளைக் கையாளும் வசதியைக் கொண்டது இது. தெலுங்கு வள மையம் **திருஷ்டி** என்ற ஒளியெழுத்துணரியை உருவாக்கியுள்ளது.

தேடி Search Engine

தேனி, யாழ் முதலான தேடிகள் தமிழுக்காக உருவாக்கப்பட்டுள்ளன. **அன்வேஷனம்** என்பது மலையாள மொழித் தேடி.

கணினிவழிக் கற்றல் CALL

கணினிவழிக் கற்கும் முறை இன்று அதிகமாக முக்கியத்துவம் பெற்று வருகிறது. இவ்வகை மென்மங்கள் குறுவட்டில் மட்டுமன்றி இணையம் மூலமாகவும் தரப்படுகின்றன. இவ்வகைக் கல்விமுறை முக்கியமான திராவிட மொழிகள் நான்கிலும் உள்ளன. கல்வியாட்டுக் கோள்கள் (Edutainment Satellite) நிறுவப்பட்ட பிறகு இந்த வலைத் தளங்களின் முக்கியத்துவம் மேலும் அதிகமாகி உள்ளது.

LI LA

லி.லா (LI LA) என்பது Learn Indian Language through Artificial Intelligence என்பதன் முதலெழுத்துச் சொல். இது சிடாக்கால் உருவாக்கப்பட்டது. தானே கற்றல் முறையில் அமைந்த மென்மம் இது. இணையத்திலும் இந்த மென்மம் கிடைக்கிறது. இந்தி எழுத்துக்களை எழுதும் பயிற்சி தரப்படுகிறது. உச்சரிப்புப் பயிற்சியும் தரப்படுகிறது. ஒருவர் படிக்கும்போது அந்தப் பகுதி வெளிச்சம்போட்டுக் காட்டப்படுகிறது. இணைய அகராதி ஒன்றும் இணைக்கப்பட்டுள்ளது. அதில் சொற்பொருள், இலக்கணக் குறிப்பு, உச்சரிப்பு முதலியன இடம்பெற்றுள்ளன. தேவைப்படும் இடங்களில் பண்பாட்டுக் குறிப்புகளும் உள்ளன. தானே கற்கும் இம்மென்மத்தில் தன் மதிப்பீடு, இடையூடகப் பயிற்சி, பாடங்களுக்கான இலக்கணப் பயிற்சி ஆகியன தரப்பட்டள்ளன. பென்சில்வேனியா மொழி மையம் (Penn Language Centre) தமிழ் கற்பிப்பதற்காக ஒரு வலைமனையை உருவாக்கியுள்ளது. இதன் பெயர் 'Tamil Smart' system என்பது. இம்மென்மம் சிங்கப்பூரில் முதன் முதலாக அறிமுகப்படுத்தப்பட்டது. தமிழை கல்வியாட்டு (Edutainment Mode) முறையில் கற்பிக்கத் தயாரிக்கப்பட்ட மென்மம் இது.

நம் நாடி என்பது சிங்கப்பூர் கல்வி அமைச்சகத்தால் பராமரிக்கப்படுகிற ஒரு வலைத் தளம். 10-12 வயதுக் குழந்தைகளுக்கான ஒரு வலைத் தளம் இது. இந்தியாவில் தமிழ் இணையப் பல்கலைக்கழகம் மழலைக் கல்வி முதல் ஆராய்ச்சி வரைக்குமான இணையதளத்தை உருவாக்கியுள்ளது. மழலைக் கல்விக்கான பாடங்கள் அயல்நாடுகளில் வாழும் தமிழருக்கு மட்டுமன்றி தமிழகக் குழந்தைகளுக்கும் மிக எளிதாகத் தமிழைக் கற்றுத் தரும் முறையில் அமைக்கப்பட்டவை. மழலைக் கல்விக்கும் உயர்நிலைக் கல்விக்கும் உதவக் கூடிய வேறு எந்தவொரு தளமும், தமிழைத் தவிர, வேறு எந்தவொரு மொழிக்கும் இன்றும் உருவாக்கப்படவில்லை எனலாம். தமிழ் கற்பிக்கும் மென்மங்களை உருவாக்கும் பணியில் **மீடியா ஃப்யூசன், டியூடர் இந்தியா பிரைவேட் லிமிடெட், சாஃப்ட் வியூ கம்ப்யூட்டர்ஸ், காட்கிராஃப்** முதலிய நிறுவனங்கள் தமிழிலே கல்வி மென்மங்களை உருவாக்கித் தந்தன.

Computer aided language learning software is seen in the website languageshome.com என்கிற இணைய தளத்திலே கணினி வழிக் கற்கும் வசதி உள்ளது. இதிலே மலையாள மொழி ஏழு பிற மொழிகளுடன் தரப்பட்டுள்ளது. இந்தப் பிற ஏழு மொழிகளின் வாயிலாக மலையாளத்தைக் கற்க முடியும். அதே போன்று மலையாளம் மூலம் அந்த ஏழு மொழிகளையும் கற்க முடியும்.

வித்யாரம்பம் என்கிற மலையாள மென்மம் ஒரு தொடக்கக் கல்விக்கான கற்பித்தல் மென்மம். குறுவட்டு வடிவில் உள்ள இம்மென்மம் மலையாள எழுத்துக்களை அறிந்துகொள்ளவும் எழுதவும் படிக்கவும் சொல்லித்தருகிறது. ஏழு பாடங்களின் மூலம் மலையாளத்தைப் படிக்க எழுத முடியும் என்று கூறப்படுகிறது. மலையாள வள மையம் **எழுத்தச்சன்** என்ற பெயரில் கற்பிக்கும் மென்மத்தை உருவாக்கியுள்ளது. பல்லாடக மென்மம் இது. உச்சரிப்புப் பயிற்சியும் உடையது. எழுத்துப் பயிற்சிக்காக, நிழல் வடிவத்தில் உள்ள எழுத்துக்களின் மீது எழுதிப்பழக வேண்டும். இம்மையத்தால் உருவாக்கப்பட்டுள்ள மற்றொரு மென்மம் **ஆசாரியன்** என்பது. ஆங்கிலத்தின் அடிப்படைப் பாடங்களை மலையாள மொழியின் மூலம் கற்பிக்கும் நோக்கத்துடன் இது தயாரிக்கப்பட்டுள்ளது. அசைவூட்டத்துடன் கூடிய பல்லாடக மென்மம். இது இடையூடக வசதி கொண்டது. **வித்யா** என்பது தெலுங்கு வள மையத்தால் உருவாக்கப்பட்டுள்ள கல்வி மென்மம்.

மின் பதிப்புகள்

தமிழிலே மின்னூல்களை அளிக்கும் பல தளங்கள் உள்ளன. **மதுரைத் திட்டம், சென்னை நூலகம்** முதலான பல தளங்கள் மின்னூல்களைத் தருகின்றன. மலையாள வள மையமும் தெலுங்கு வள மையமும் மின்னூல்கள் பலவற்றை இணையத்தில் தருகின்றன. மலையாள நூல்களின் அடைவுகளும் இணையத்தில் தரப்பட்டு வருகின்றன. தமிழிலே மிக அதிகமான மின்னிதழ்கள் இணையத்தில் இடம்பெறுகின்றன. தமிழ் மரபு அறக்கட்டளை தமிழிலுள்ள அரிய நூல்களையும் சுவடிகளையும் மின்பதிப்பாக வெளியிட்டு வருகிறது. தமிழ் இணையப் பல்கலைக்கழக தளத்திலும் ஏட்டுச் சுவடிகள், நூல்கள் ஆகியவற்றின் மின்பதிப்புகளைக் காணமுடிகிறது. மின் பதிப்புகளைப் பொறுத்தவரையில் தமிழிலே மிக அதிகமான நூல்கள் மின்பதிப்பாக்கம் பெற்றுள்ளதை அறிய முடிகிறது.

திராவிட மொழிகளில் கணினி முயற்சிகள் நடைபெற்றுவருகின்றன. இம்முயற்சிகளில் வளமையங்கள், சிடாக், பாஷா இந்தியா, இந்தியத் தொழில்நுட்ப நிறுவனங்கள், தனியார் நிறுவனங்கள், தனிப்பட்ட ஆர்வலர்கள் என பலருடைய பங்களிப்பும் உள்ளது. கணினியின் தொடக்க காலகட்டத்தில் எழுதிகளையும் உரைப்பதிப்பிகளையும் உருவாக்கிய இம்மொழியினர் தற்போது மின்னூல்கள், பேச்சுணரிகள், எழுத்துணரிகள், ஒளியெழுத்துணரிகள், எந்திர மொழிபெயர்ப்பு, உருப்பு பகுப்பாய்வு, தரவுக்களத் தயாரிப்பு எனப் பல்வேறு பணிகளில் தம்மை ஈடுபடுத்திக் கொண்டுள்ளனர். யுனிகோடு குறியீட்டில் திராவிட மொழிகள் பல எழுத்துருக்களையும் உருவாக்கியுள்ளன. மேலும் சமூக அக்கறையுடன் நம் மொழி-கணினி வல்லுநர்கள் பல பணிகளையும் செய்து வருகின்றனர். மாற்றுத் திறனுடையோரில் பார்வைக் குறைபாடுடையோருக்கான மென்மத் தயாரிப்பும் நடைபெற்று வருகிறது. இன்னும் கேட்கும் திறனும் பேசும் திறனும் அற்றோருக்கான மென்மத் தயாரிப்புகள் நம் மொழிகளிலோ பிற இந்திய மொழிகளிலோ நடைபெற்றதாகத் தெரியவில்லை. அவை பற்றியும் நம் திராவிட மொழி வல்லுநர்களும் கணினி அறிஞர்களும் சிந்திக்க வேண்டும்.

Conceptual Lexicon for Knowledge Representation

S.Rajendran
Tamil University
Thanjavur
raj_ushush@yahoo.com

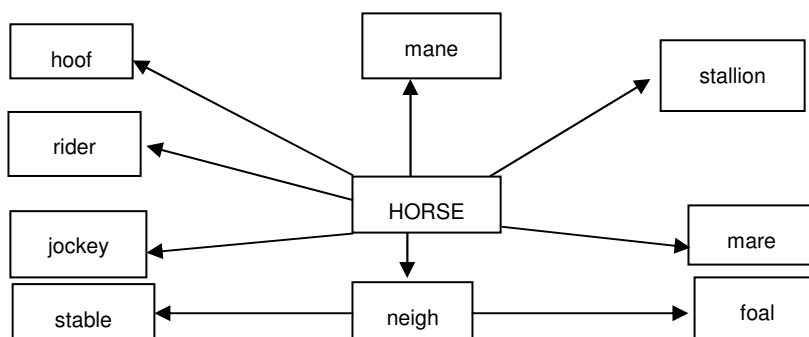
Introduction

Conceptual graphs emphasize semantics. The earliest forms, called existential graphs, were invented by the philosopher Charles Sander Peirce (1897) as a graphical notation for symbolic logic. Lucien Teniere (1959) used similar graphs for his dependency grammar. The earliest form implemented on a computer was the correlational nets by Silvio Cecato (1961), who used them as intermediate language for machine translation. There are philosophical and psychological evidence that conceptual graphs are mental representation unbounded by knowledge of a particular language.

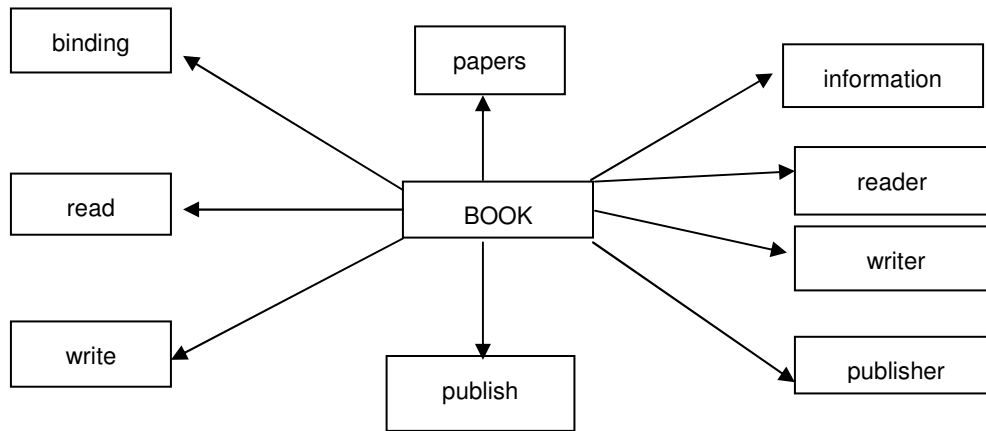
The proposed conceptual lexicon has concepts as its entries which are independent of a specific language and the meanings of concepts are given in terms of conceptual graphs from which the surface representation of lexical items belonging to a particular language can be derived. The proposed lexicon can be manipulated to generate a text in the form of a target language. The theory propounded by Sowa (1984) has been exploited to suit our purpose. At the same time the four levels of representations proposed for the a generative lexicon (Pustejovsky, 1995) and the semantic representation in WordNet (Pike Vasson, 2000) are also kept in mind while writing the meaning of a lexical item by means of conceptual graph.

Why Conceptual Lexicon

Dictionary definitions are mostly inadequate representations of words or concepts. A real definition will become encyclopedic. Let us take the concepts "horse" and "book". "Horse" may require at least the following representation:



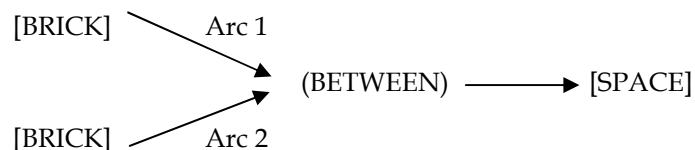
Similarly “book” requires the following representation:



For a universal representation of concepts across languages conceptual graphs may be used.

Conceptual Graphs

Conceptual graphs from a knowledge representation language based on linguistics, psychology and philosophy (Sowa, 1984: 69). Concepts are language independent ones derived from percepts. A conceptual graph is a finite, connected, bipartite graph. The two kinds of nodes of the bipartite graph are concepts and conceptual relations. Every conceptual relation has one or more arcs, each of which must be linked to some concept. If a relation has n arcs, it is said to be n -adic, and its arcs are labeled 1, 2, ..., n . The term monadic is synonymous with 1-adic, dyadic with 2-adic and triadic with 3-adic. A single concept by itself may form a conceptual graph, but every arc of every conceptual relation must be linked to some concept. Concepts are discrete units. Combinations of concepts are not diffuse mixtures, but ordered structures. Only discrete relations are recorded in concepts. Continuous forms must be approximated by patterns of discrete units. For example, *a space between a brick and a brick* can be represented as follows (Sowa, 1984:72):



In the graphs, concept nodes represent entities, attributes, states, and events, and relation nodes show how the concepts are interconnected. Distinctions can be made between simple and complex concepts. Simple concepts are basic concepts from which complex concepts can be derived.

Semantic Net Work

Although the concept types CAT and TOMATO map directly to percepts, other types like PRICE, FUNCTION and JUSTICE have no sensory correlates. Abstract concepts acquire their meanings not

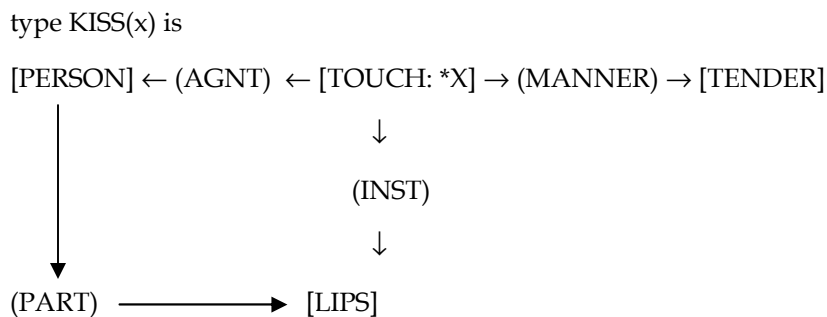
through direct associations with percepts, but through a vast network of relationships that ultimately links them to concrete concepts. A conceptual graph has no meaning in isolation. For example, the description of the concept, MAN is represented as follows:

[MAN] → (ISA) → [HUMAN BEING] → (ISA) → [ANIMAL]

Abstraction and Definition

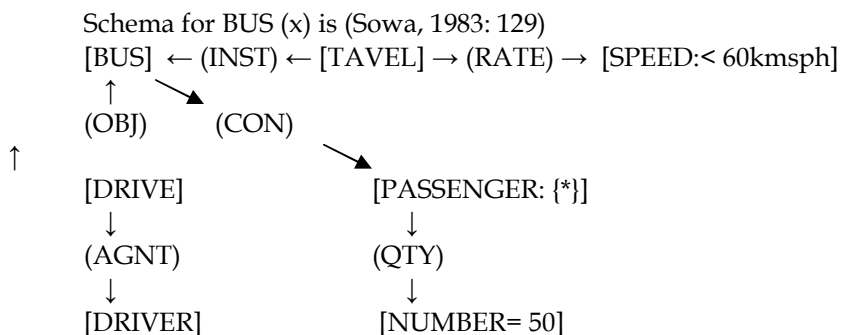
Definition can specify a type in two different ways: by stating necessary and sufficient conditions for the type, or by giving a few examples and saying that everything similar to these belongs to the type. The first method derives from Aristotle's method of definition by genus and differentiae. And the second method is closer to Wittgenstein (1953). AI systems have supported both methods. Conceptual graphs support type definitions by genus and differentiae as well as schemata and prototypes.

Type definition for KISS (Sowa, 1984:106)



Schemata

The basic structure for representing background knowledge for human-like inference is called the schemata. It is a pattern derived from past experience that is used for interpreting, planning, and imagining other experiences. Schemata incorporate domain-specific knowledge about the typical constellations of entities, attributes and events in the real world. Schemata are similar in structure to type definition. Yet concept type may have at most one definition, but arbitrarily many schemata. Type definitions present the narrow notion of a concept, and schemata present the broad notion. Type definitions are obligatory conditions that state only the essential properties, but schemata are optional defaults that state the commonly associated accidental properties.



Schemata show the typical ways in which a concept may be used, but they do not describe a typical instance of a concept.

PROTOTYPE

A prototype is a typical instance. Instead of describing a specific individual, it describes a typical of “average individual”. A Schema for ELEPHANT might specify a range of characteristics for elephants or a range of behaviours and habitats for elephants. A prototype ELEPHANT would combine and restrict such schema to describe a typical elephant.

Proto type for ELEPHANT (x) is (Sowa 1984: 136):

[ELEPHANT:*x] -

(CHAC) → [HEIGHT: @ 3.3 m]

(CHAC) → [WEIGHT: @ 5400 kg]

(COLR) → [DARK-GREY]

(PART) → [NOSE]-

(ATTR) → [PREHENSILE]

(IDNT) → [TRUNK]

(PART) → [EAR] {*} -

(QTY) → [NUMBER:2]

(ATTR) → [FLOPPY]

(PART) → [TUSK: {*}] -

(QTY) → [NUMBER:2]

(MATR) → [IVORY]

(PART) → [LEG: {*}]

(QTY) → [NUMBER:4]

(STAT) → [LIVE] -

(LOC) → [CONTINENT: {Africa | Asia}]

(DUR) → [TIME: @ 50 YEARS]

Conceptual Representation

Some of the conceptual relations listed in Sowa (1984) are adopted to suit our purpose.

- accompaniment. (ACCM) links [ENTITY:*x] to [ENTITY:*y] where *y is accompanying *x.
- agent. [AGNT] links [ACT] to [ANIMATE], where ANIMATE concept represents the actor of the action.
- attribute. (ATTR) links [ENTITY:*x] to [ENTITY:*y] where *x has an attribute *y.
- cause. (CASE) links [STATE:*x] to [STATE:*y] where *x has a cause *y.
- characteristic. (CHRC) links [ENTITY:*x] to [ENTITY]
- destination. (DEST) links [ACT] to [ENTITY] towards which the action is directed.
- experience (EXPR) links [STATE] to [ANIMATE], who is experiencing that state.
- instrument. (INST) links [ENITY] to [ACT] in which the entity is causally involved.

The following relations are also taken in to account to define a concept using some other concept.

Hypernymy- Hyponymy		ANIMAL → MAMMAL
Hyponymy- Hypernymy		COW → MAMMAL
Holonymy-Meronymy	Wholes to parts	TABLE → LEG
“	Groups to members	DEPARTMENT → PROFESSOR
Meronymy-Holonymy	Parts to wholes	WHEEL → CART
Troponymy	From events to their subtypes	WALK → LIMP
Entailment	From events to the events they entail	SNORE → SLEEP

Pustejovsky (2001:56) characterize a generative lexicon as a computational system involving at least the following levels of representation:

1. ARGUMENT STRUCTURE: Specification of number and type of logical arguments
2. EVENT STRUCTURE: Definition of the event type of an expression and its subeventual structure
3. QUALIA STRUCTURE: A structural differentiation of the predicative force for a lexical item
4. LEXICAL INHERITANCE STRUCTURE: Identification of how a lexical structure is related to other structures in this type of lattice

Pustejovsky (2001:56) assumes that word meaning is structured on the basis of four generative factors, or qualia roles, that capture how humans understand objects and relations in the world and provide the minimal explanation for the linguistic behaviour of lexical items.

CONSTITUTIVE: the relation between an object and its constituent parts

FORMAL: the basic category that distinguishes the object within a larger domain

TELIC: the object's purpose and function

AGENTIVE: factors involved in the object's origin or "coming into being."

The qualia structure is the core of the generative properties of the lexicon, because it provides a general strategy for creating increasingly specific concepts with conjunctive properties. A simple schematic description of a lexical item, α , using this representation is shown below:

α
 ARGSTR = ARG1=x
 ...
 CONST = what x is made of
 QUALIA = FORMAL = what x is
 TELIC = function of x
 AGENTIVE = how x come into being

The lexical structure for *book* as an object can then be represented as follows:

book

ARG1 = y:information
ARGSTR = ARG2 = x:phys_obj
 information.phy_obj
FORM = holds (x,y)
QUALIA = TELIC = read (e,w,x,y)
AGENT = write (e', v, x, y)

The ideas propounded by Pustejovsky will also be taken into consideration while defining a concept.

The following is the sample of the conceptual lexicon. Each item is a concept and the concepts will be mapped against the lexical items of a language.

[CAT] → (ISA) → [ANIMAL]
[PENCIL] → (ISA) → [INSTRUMENT]
 ↓
 (FUNCT)
 ↓
 [WRITING]

The verbs are provided with argument structures. A frame of arguments will be given with their necessary relations. The verbal concept ACT represented in the following fashion.

[ACT] → (ISA) → [EVENT]-
 (AGENT) → [ANIMATE ENTITY]
[ARRIVE] → (ISA) → [EVENT] -
 (AGENT) → [MOBILE-ENTITY]
 (GOAL) → [LOCATION]

Lexical and Conceptual Structures

Each natural language has a well-organized lexical and syntactic system. Each domain of knowledge has a well-organized conceptual system. Complexities arise because each language tends to use and reuse the same words and lexical patterns in many different conceptual domains. The lexical structures are

- Relatively domain independent,
- Dependent on syntax and word forms,
- Highly language dependent.

And the conceptual structures are

- Highly domain dependent,
- Independent of syntax and word forms,
- Language independent, but possibly culture dependent.

When there are cross-linguistic similarities in lexical patterns, they usually result from underlying conceptual similarities. English verb *give*, for example, takes a subject, object, and indirect object. Other languages may have different cases marked by different prepositions, postpositions, inflections, and word order; but the verb that mean roughly the same as *give* also have three participants – a giver, a thing given, and a recipient. In all languages, the three participants in the conceptual pattern lead to three arguments in the lexical patterns.

The distinction between lexical structures and conceptual structures addresses the following things:

- Lexical structures are oriented towards language. The representation developed here is strongly influenced by linguistic theories of syntax and thematic roles.
- Conceptual structures are designed for representing knowledge about the world. They may grow too large to be expressed in a single sentence, and they may contain concepts types that cannot be expressed by a single word.
- Since they can be represented by similar structures, the same operations can be used on them. Furthermore, lexical structures can be converted to deeper conceptual structures by a step-by-step process, not by a translation between radically different forms.
- Finally, common structures facilitate language learning and conceptual creativity. In learning, a child generalizes conceptual structures learned from experience to form the initial lexical structures needed for language. Metaphor and conceptual refinement create new conceptual structures by adapting old lexical structures to novel situations.

Conclusion

The distinction between lexical structures and conceptual structures provides a principled basis for partitioning knowledge into the lexicon and the more detailed knowledge about the world. Conceptual graphs provide a formalism for representing both kinds of structures with a level of precision that allows deeper and more systematic analysis of the relationship between them. As a result, they can help to replace vague discussion with a precise methodology that has a greater chance of being computerized. Finally the direct mapping between conceptual graph and natural language can simplify the task of knowledge acquisition: a knowledge base of conceptual graphs could be generated directly from natural language inputs. After being primed with a dictionary of lexical knowledge, the system could build up its own encyclopaedia of world with the aid of a tutor communicating in English, not a knowledge engineer coding in a specialized notation.

Bibliography

1. Bouillon, P and Busa, F. 2001 "Qualia and the Structuring of Verb Meaning." In: P. Bouillon and F. Busa (eds.). *The Language of Word Meaning*. Cambridge: Cambridge University Press, 149-167.
2. -----2001. "Type Construction and the Logic of Concepts." In: P. Bouillon and F. Busa (eds). *Language of Word Meaning*. Cambridge: Cambridge University Press, 91-123.
3. Busa, F & Calzolari, N and Lenci, A. 2001. "Generative Lexicon and the SIMPLE Model: Developing Semantic Resources for NLP." In: P. Bouillon and F. Busa (eds). *Language of Word Meaning*. Cambridge: Cambridge University Press, 333-349.
4. Cruse, D.A. 1986. *Lexical Semantics*. New York: Cambridge University Press.
5. Grimshaw, J. 1990. *Argument Structure*. Cambridge: MIT Press.
6. Fodor, J.A. and Lepore, E. 2001. "The Emptiness of the Lexicon: Critical Reflections on J. Pustejovsky's "The Generative Lexicon." In: P. Bouillon and F. Busa (eds). *Language of Word Meaning*. Cambridge: Cambridge University Press, 28-49.
7. McGilvary, J. 2001. "Chomsky on the Creative Aspect of Language Use and Its Implications for Lexical Semantic Studies." In: P. Bouillon and F. Busa (eds). *Language of Word Meaning*. Cambridge: Cambridge University Press, 5-27.

8. Pustejovsky, J. 1994. "A Richer Characterization of Dictionary Entries: The Role of Knowledge Representation." In B.T.S. Atkins, and A. Zampolli (eds.). *Computational Approach to Lexicon*. Oxford: Oxford University Press.
9. -----1995a. "Linguistic Constraint on Type Coercion." P. Saint-Dizier and E. Viegas (eds.). *Computational Lexical Semantics*. Cambridge: Cambridge University Press.
10. -----1995b. *The Generative Lexicon*. Cambridge, Massachusetts: The MIT Press.
11. -----2001. "Generativity and Explanation in Semantics: A Reply to Fodor and Lepore." In *Language of Word Meaning*. Cambridge: Cambridge University Press, 51-74.
12. Peirce, C.S. 1960. *Collected papers of Charles Sanders Peirce*. Peirce. Arthur W. Burks(ed.) 8 vols., Cambridge: Harvard University Press.
13. Pustejovsky, J. and Boguraev, B. 1993. "Lexical Knowledge Representation and Natural Language Processing." In: F. Pereira and B. Grosz (eds.), *Natural Language Processing*. Cambridge, Mass.: MIT Press.
14. Rajendran, S. 1978. *Syntax and Semantics of Tamil Verbs* (manuscript). Ph.D. Thesis. Poona: University of Poona.
15. -----1983. *Semantics of Tamil Vocabulary*. (Report of the UGC sponsored Postdoctoral Work in manuscript). Poona: Deccan College Post Graduate and Research Institute.
16. ----- 2002. "Preliminaries to the Preparation of a Word Net for Tamil." *Language in India* 2:1, www.langueinindia.com
17. ----- 2003. "Creating Generative Lexicon from Dictionaries: Tamil Experience." In: *Recent Advances in Natural Language Processing: Proceedings of the ICON 2003*. Mysore: CIIL, 83-91.
18. -----2004. *Priorities in the Pursuit of Preparing a Generative Lexicon for Tamil*. Paper read in ICOIL held in CASL, Annamalai University.
19. Rajendran, S., Arulmozi S., Kumara Shanmugam, B., Baskaran, S. and Thiyagarajan, S. 2002. "Tamil WordNet." In *Proceedings of the First International Global WordNet Conference*. Mysore: CIIL, 271-274.
20. Ruimy, N. Gola, E. and Monachini, M. 2001. *Lexicography Informs Lexical Semantics: The SIMPLE Experience*. In P. Boulillon and F. Busa, (eds). *Language of Word Meaning*. Cambridge: Cambridge University Press.
21. Saint-Dizier, P and Viegas, E. (eds.) 1995. *Computational Lexical Semantics*. Cambridge: Cambridge University Press.
22. Somers, H.L. 1987. *Valency and Case in Computational Linguistics*. Edinburgh: Edinburgh University Press.
23. Sowa, J.F. 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Reading: Addison-Wesley Publishing Company.
24. ----- 1988. "Using a Lexicon of Canonical Graphs in a Semantic Interpreter." In M. Evens (ed.) *Relational Models of the Lexicon*. New York: Cambridge University Press, pp 73-97.
25. ----- 1993. "Lexical Structure and Conceptual Structures." In: *Semantics and the Lexicon*.
26. Dordrecht: Kluwer Academic Publishers.
27. Sowa, J.F. and Eileen, C.W. 1986. "Implementing a Semantic Interpreter for Conceptual Graphs." *IBM Research and Development* 30 (1), 57-69.

கணினித்தமிழ் வளர்ச்சிக்கான அடிப்படை ஆராய்ச்சிப் பணிகள்

ந. தெய்வ சுந்தரம்

பேராசிரியர் & தலைவர்

தமிழ்மொழித்துறை,

சென்னைப் பல்கலைக்கழகம்

ndsundaram@hotmail.com

கணினித்தமிழ் வளர்ச்சி என்பது கணினியானது தமிழ்த்தொடர்களைப் புரிந்துகொள்ளவும் (Understanding) அவற்றை உருவாக்கவும் (Generation) தேவையான தமிழ்மொழி அறிவை அதற்குக் கொடுப்பதற்காக நாம் மேற்கொள்ளவேண்டிய பணிகளைக் குறித்து நிற்கிறது.

கருத்துப்புலப்படுத்தும் இயற்கைமொழிகளும் (Communication and Natural Languages)

மனித சமுதாயத்தில் பொருள் உற்பத்தி மற்றும் பிற சமூக நடவடிக்கைகளில் மனிதர்களுக்கிடையே கருத்துப்புலப்படுத்தம் என்பது மிக அடிப்படையான ஒன்றாகும். நமது கருத்துப்புலப்படுத்தச் செயற்பாட்டில் இயற்கைமொழிகள் இன்றியமையாதவையாக அமைகின்றன. மொழிசாராக் கூறுகளான (Non-Verbal Means) முகபாவம், உடல் அசைவு, படம், அட்டவணை போன்ற பலவற்றையும் நாம் இயற்கைமொழியோடு இணைத்து, கருத்துப்புலப்படுத்தத்தில் ஈடுபடுகிறோம்.

இயற்கைமொழி அறிவுத்திறனானது மனிதருக்கே - மனித மூளைக்கே - உரிய ஒன்றாகும் (Human Species-specific). மனிதரையும் பிற உயிரினங்களையும் வேறுபடுத்தக்கூடிய முக்கியமான இரண்டு பண்புகளில் ஒன்று மொழி, மற்றொன்று உழைப்பு. நமது மூளையில் இயற்கைமொழி அறிவு எப்படி உருவாகிறது, அம்மொழி அறிவின் கூறுகள் எவை, அவற்றின் அமைப்பு என்ன, அந்த அறிவு எங்கே , எப்படி தேக்கிவைக்கப்படுகிறது, எவ்வாறு பயன்படுத்தப்படுகிறது என்பதை ஆராயும் அடிப்படை அறிவியலே மொழியியலாகும் (Linguistics). கடந்த ஐம்பது ஆண்டுகளில் அமெரிக்க மொழியியல் அறிஞர் நோம் சாம்ஸ்கியின் (Noam Chomsky) வழிகாட்டுதலில் நடைபெற்றுவரும் மாற்றிலக்கண மொழியியல் ஆய்வானது (Generative Grammar) மேற்கூறிய துறையில் குறிப்பிடத்தக்க முன்னேற்றத்தைக் கண்டுள்ளது. மொழியியலின் இப்பிரிவானது (புலன்) மனித அறிவாற்றல் இயலின் (Science of Human Cognition) ஒரு பகுதியாக அமைகிறது.

இயற்கைமொழி ஆய்வு (Natural Language Processing - NLP)

இயற்கைமொழி அறிவுத்திறன் பற்றிய மேற்குறிப்பிட்ட மொழியியலின் ஆய்வுகளைச் சோதித்தறிய (to verify) தற்போது கணினிப்பொறி மிகவும் பயன்படுகிறது. இயற்கைமொழி அறிவுத்திறன் பற்றிய குறிப்பிட்ட கோட்பாடுகள் மற்றும் மாதிரிவடிவங்களைச் (Principles and Models) சோதித்தறிய வேண்டும் என்றால், அவற்றின் அடிப்படையில் ஒரு குறிப்பிட்ட மொழியின் அறிவை ஒரு மின்னணு மாதிரிவடிவத்தில் (Model of Electronic Representation) கணினியில் அமைத்துச் சோதிக்கலாம்.

இதன் அடுத்த கட்டம், நமது கருத்துப்புலப்படுத்தத் திறனை வெளிப்படுத்தும் ஒரு மின்னணு மாதிரிவடிவத்தைக் (Electronic Model for Human Communicative Action) கணினியில் உருவாக்கி, கணினியை நம்மைப் போன்றே இயற்கைமொழி அறிவு மற்றும் பிற பின்புல உலக அறிவை (Pragmatic and

other World knowledge) அடிப்படையாகக்கொண்டு, கருத்துப்பலப்படுத்தச் செயற்பாட்டில் - கருத்தாடலில் (Discourse) - ஈடுபட வைக்கும் முயற்சியாகும். இந்த ஆய்வின் மற்றொரு பயன்பாடு, கணினிக்குச் செயற்கை அறிவைக் (Artificial Intelligence - AI) கொடுக்கும் முயற்சியின் ஒரு பகுதியாகவும் இது அமைவதாகும். மனித மூளையின் அறிவுத்திறனில் ஒரு பகுதியான தர்க்கவியல் கூறுகள் (Logical Knowledge) இயற்கைமொழிகளின் அமைப்பில் புதைந்து நிற்பதால் (Embedded), இயற்கைமொழிகளைக் கையாளும் ஆற்றலைக் கணினிக்குக் கொடுக்கும் பணியானது செயற்கை அறிவாற்றல் அறிவியலின் (Science of Artificial Intelligence) ஒரு பகுதியாகவும் அமைகிறது. இவ்வாறு கணினிக்கு இயற்கைமொழிகளைக் கையாளும் திறனை அளிக்கும் துறையே இயற்கைமொழி ஆய்வு என வழங்கப்படுகிறது.

கணினிமொழியியல் (Computational Linguistics)

கணினிக்கு இயற்கைமொழி அறிவை எவ்வாறு கொடுப்பது (Natural Language Knowledge Representation) என்பது நம் முன் நிற்கும் அடுத்த வினா. 0, 1 என்ற இரட்டை இலக்குகளையும் (Binary Numbers) / எண்களையும் அதன் அடிப்படையிலான தர்க்கவியலையும் - அடிப்படையாகக்கொண்டு செயல்படும் கணினிக்கு இயற்கைமொழி அறிவைக் கொடுப்பதற்குக் கணினியியல், கணிதவியல், புள்ளியியல், மின்னணுவியல் ஆகிய அறிவியல் துறைகளைச் சார்ந்த அறிவு தேவைப்படுகிறது. இத்துறை அறிவுகளை அடிப்படையாகக்கொண்டு, இயற்கைமொழி அறிவைக் கணினிக்கேற்ற வகையில் தரவுத்தளம் (Data Structure) மற்றும் கணினிநிரல்வழிமுறைகள் (Algorithms) ஆகியவைகளாக மாற்றி, மின்னணு வடிவத்தில் கணினிக்குக் கொடுக்கவேண்டியுள்ளது. அதாவது இயற்கைமொழி அமைப்பை - இலக்கணத்தை - கணினிக்கேற்ற மின்னணு இலக்கணமாக (Electronic Grammar) மாற்றிக் கொடுக்கவேண்டியுள்ளது. இப்பணியை மேற்கொள்ளும் துறையே கணினிமொழியியல் என்றழைக்கப்படுகிறது.

மொழித்தொழில்நுட்பம் (Language Technology)

கணினிமொழியியலின் துணைகொண்டு, கணினிக்கு இயற்கைமொழி ஆய்வை மேற்கொள்ளும் அறிவுத்திறனைக் கொடுப்பதன்மூலம் பல மொழிவழிச்செயல்களை உள்ளடக்கிய பணிகளை (Language Activities) மேற்கொள்ளும் திறனையும் கணினிக்குக் கொடுக்க இயலுகிறது. ஒரு எழுத்து உரையை (Written Text) மின்னணு உரையாக மாற்றுவது (Digitalization / Creating Electronic Text) , ஒரு பேச்சு உரையை (Spoken Text) மின்னணு உரையாக மாற்றும் பணி (Automatic Speech Recognizer - ASR) , உருவாக்கப்பட்ட மின்னணு உரையைச் சரிபார்த்துப் பதிப்பிக்கும் பணி (Word Processing), மின்னணு உரையைப் பேச்சு உரையாக மாற்றும் பணி (Text to Speech - TTS) , மின்னணு உரையில் தேவையான தகவல்களைத் தேடித்தரும் பணி (Information Extraction and Retrieval - IE & IR) , குறிப்பிட்ட சொற்கள் அல்லது தொடர்களை உள்ளடக்கிய மின்னணு உரைகளை இணையத்தில் தேடித்தரும் தேடல் பொறி (Search Engine) இவையெல்லாம் மேற்கூறிய மொழிவழிச்செயல்களை உள்ளடக்கிய மின்னணு மொழிக்கருவிகளுக்கு (Electronic Language Tools) எடுத்துக்காட்டுகளாகும். இதுபோன்ற மின்னணு மொழிக்கருவிகளை உருவாக்கும் துறையே மொழித்தொழில்நுட்பம் (Language Technology) என்றழைக்கப்படுகிறது. எழுத்துரு உருவாக்கம் (Font Development) , மின்னணு விசைப்பலகை உருவாக்கம் (Electronic Keyboard) போன்ற அடிப்படைப் பணிகளில் இருந்து, தானியங்கி மொழிபெயர்ப்பு (Automatic Machine Translation - MT) , மனிதன் - கணினி கருத்துப்பறிமாற்றம் (Human - Machine Interface) போன்ற உயர்மட்டப் பணிகள்வரை கணினிவழி செய்ய முயலுவதே இத்துறையின் நோக்கமாக அமைகிறது.

மின்னணு மொழிவளர்ச்சித்திட்டம் (E-language Planning - ELP)

இயற்கைமொழி ஆய்வு, கணினிமொழியியல், மொழித்தொழில்நுட்பம் என்ற மேற்குறிப்பிட்ட மூன்று துறைகளின் பயன்பாட்டையும் தமிழுக்குக் கொண்டு வந்து சேர்ப்பது இன்றைய உலகமயமாக்கச் சூழலில் (Globalization) மிக முக்கியமான ஒரு மொழிவளர்ச்சிப் பணியாகும். செயற்படுத்த மொழியியலின் (Applied Linguistics) ஒரு பிரிவான மொழிவளர்ச்சித்திட்டத்தில் (Language Planning) எழுத்துருவாக்கம் (Graphization), புதுமையாக்கம் (Modernization), தரப்படுத்தம் (Standardization) போன்ற பல்வேறு மொழிவளர்ச்சித் திட்டப் பணிகள் மேற்கொள்ளப்படுகிறது. அத்தோடு இன்றைய அறிவியல் உலகில் ஒரு மொழியை மின்னணு மொழிக்கு (Electronic Language) உரிய தகுதியை உருவாக்குவதும் ஒரு இன்றியமையாத பணியாகும். இப்பணியை மின்னணு மொழிவளர்ச்சித்திட்டம் என்றழைக்கலாம்.

உலகமயமாக்கச் சூழலில் தகவல்தொழில்நுட்ப (Information Technology) வளர்ச்சியின் முக்கியத்துவம் அனைவருக்கும் தெரிந்ததே. காலமும் இடைவெளியும் தகவல் பரிமாற்றத்திற்கு இன்று ஒரு தடையே இல்லை. வினாடிகளில் பத்தாயிரக்கணக்கான மைல்களுக்கு அப்பால் தகவல்களைப் பரிமாறிக்கொள்ள வாய்ப்புகளை இன்றைய அறிவியல் , தொழில்நுட்ப வளர்ச்சி நமக்கு அளித்துள்ளது. கணினி, கைபேசி போன்ற தகவல்தொழில்நுட்பக்கருவிகளின் வளர்ச்சி வியப்புக்குரியதாக இன்று அமைந்துள்ளது. இக்கருவிகளில் எந்தெந்த மொழிகள் இடம் பெறுகின்றனவோ அந்த மொழிகளே உலக முக்கியத்துவம் வாய்ந்த மொழிகளாக இனி விளங்கமுடியும். மேலும் தகவல்தொழில்நுட்பத்திற்கான கணினி போன்ற மின்னணுக் கருவிகளின் வழியே கிடைக்கும் பயன்பாடுகளை - அறிவுத் திரட்டல்களை- அனைத்து மக்களுக்கும் கிடைக்க வழிசெய்யவேண்டும் என்றால், அக்கருவிகளில் அம்மக்களின் மொழிகள் இடம் பெறவேண்டும். இல்லையெனில் அது மின்னணுக் கருவியினால் ஏற்படும் பிளவை (Digital Divide) சமுதாயத்தில் ஏற்படுத்தும்.

மேற்குறிப்பிட்ட கருத்துகளின் அடிப்படையில் தமிழ்மொழியைத் தகவல் தொழில்நுட்பக் கருவிகளுக்கேற்ற ஒரு மொழியாக மாற்றவேண்டும். குறிப்பாக, கணினிக்கேற்ற ஒரு மொழியாக வளர்த்தெடுக்கவேண்டும். இங்கு நாம் கவனத்தில் கொள்ளவேண்டிய ஒன்று, கணினிக்காகத் தமிழ் அமைப்பை மாற்றுவதோ அல்லது ஒரு புதுவகைத் தமிழ்வழக்கை உருவாக்குவதோ அல்ல இப்பணி. மாறாக, நிலவும் தமிழ் அமைப்பை - அறிவை- கணினிக்குப் புரியக்கூடிய வகையில் தரவுத்தளமாகவும், நிரல்வழிமுறைகளாகவும் மாற்றிக் கொடுப்பதேயாகும். இதுவே இக்கட்டுரையில் கணினித்தமிழ் வளர்ச்சி என்று அழைக்கப்படுகிறது.

தமிழ்மொழியைக் கணினிமொழியியல் அடிப்படையில் கணினித்தமிழாக வளர்த்தெடுக்க, தமிழ் ஆய்வாளர்கள் மேற்கொள்ளவேண்டிய பணிகள் பற்றி இக்கட்டுரை சில கருத்துகளை முன்வைக்கிறது.

உள்ளீடு, எழுத்துரு, விசைப்பலகை (Encoding, Font, Keyboard)

கணினியில் தமிழை உள்ளீடு செய்வதில் இருந்த சிக்கல்கள் இன்று ஒருங்குறி வருகையால் தீர்க்கப்பட்டுள்ளது. ஒருங்குறி உள்ளீட்டை அடிப்படையாகக்கொண்டு, தமிழ் மின்னணு உரையை உருவாக்கிப் பதிப்பதிலும் சில பிரச்சனைகள் இன்னும் நீடிக்கின்றன. அவற்றைத் தீர்ப்பதற்கான முயற்சிகளில் தனிநிலையில் ஆய்வாளர்களும், உத்தமம் , கணித்தமிழ்ச்சங்கம் போன்ற அமைப்புகளும், தமிழக அரசு மற்றும் ஆய்வுநிறுவனங்களும் பல நிலைகளில் முயற்சிகள் மேற்கொண்டுவருகின்றன. ஒருங்குறி பற்றிய ஒரு ஆய்வு முடிவை உத்தமம் அமைப்பின் சார்பாக ஒருங்கிணைவு முன்வீடு (Convergence Proposal) என்ற ஆவணம் தற்போது தயாரிக்கப்பட்டுள்ளது ஒரு வரவேற்கத்தக்க முயற்சியாகும். ஒருங்குறி அடிப்படையிலான பலவகை எழுத்துருகள் இன்று உருவாக்கப்பட்டுள்ளன.

மின்னணு விசைப்பலகை உருவாக்கத்திலும் நல்ல முன்னேற்றங்கள் ஏற்பட்டுள்ளன. தமிழ்இணையம் 99, தமிழ்த்தட்டச்சு , ரோமன் தட்டச்சு என்ற மூன்று விசைப்பலகைகள் தற்போது நடைமுறையில் பயன்பட்டு வருகின்றன. ரோமன் தட்டச்சுப் பலகையில் சில பிரச்சனைகள் உள்ளன. அவற்றை விரைவில்

தீர்த்துவிடலாம். இவ்வாறு உள்ளீடு, மின்னணு விசைப்பலகை ஆகிய இரண்டு அடிப்படைப் பணிகளிலும் குறிப்பிடத்தக்க வளர்ச்சியைக் கணினித்தமிழ் பெற்றுள்ளது. கைபேசி விசைப்பலகை உருவாக்கத்திலும் பலர் ஈடுபட்டுள்ளனர்.

இரட்டைவழக்குத் தமிழும் கணினிமொழியியலும் (Diglossic Tamil and Computational Linguistics)

மேற்கண்ட வளர்ச்சிகள் தமிழ்மொழியைக் கணினியில் உள்ளீடு செய்ய வழிவகுத்துள்ளது. அடுத்த கட்ட பணிகள் தமிழ்மொழியியலை அடிப்படையாகக் கொண்ட பணிகளாகும். தமிழ்மொழியை - பேச்சு வழக்கு, எழுத்து வழக்கு இரண்டையுமே - கணினி புரிந்துகொள்ளவேண்டும். தமிழ்மொழி ஒரு இரட்டை வழக்கு மொழி என்பதை இங்கு நாம் மனதில் கொள்ளவேண்டும். தமிழ்ச்சமுதாயத்தில் தமிழ்வழிச் செயல்பாடுகள் அனைத்தையும் மேற்கொள்ள இரண்டு வழக்குகளுமே தேவைப்படுகிறது. பேச்சு வழக்கில் பல வட்டார வழக்குப் பிரிவுகள் உண்டு. அதுபோன்று எழுத்துவழக்கிலும் சில வகைகள் உண்டு. இருப்பினும் அனைவரும் புரிந்துகொள்ளக்கூடிய ஊடக எழுத்துத்தமிழைத் தொடக்கத்தில் கணினி எழுத்துத்தமிழாக எடுத்துக்கொள்ளலாம். பேச்சு வழக்கைப் பொறுத்தவரையில் இன்று கல்வி, போக்குவரத்து ஆகியவற்றின் வளர்ச்சிகளின் பயனாக ஒரு பொதுப்பேச்சுத் தமிழ் உருவாகியுள்ளது. இந்த வழக்கு (பிராமணர் அல்லாத) முறைசார் கல்வியைப் பெற்றவர்களின் பேச்சுவழக்காக அமைந்துள்ளது என்று தமிழ்மொழியியலாளர்கள் கருதுகின்றனர். இந்தப் பொதுப்பேச்சுத்தமிழைக் கணினிப்பயன்பாட்டுக்கு எடுத்துக்கொள்ளலாம்.

மேற்கண்ட இரு தமிழ் வழக்குகளையும் உள்ளடக்கிய தமிழைக் கணினி புரிந்துகொள்ளவேண்டும். இதற்கு நாம் மேற்கொள்ளவேண்டிய பணிகளை இனி பார்க்கலாம்.

பொருள் மயக்கம் (Sense Disambiguation)

மொழியில் நாம் உருவாக்கும் தொடர்கள் , நாம் புலப்படுத்த விரும்புகிற கருத்தை அல்லது பொருளை அடையாளப்படுத்தி (Representing) நிற்கின்றனவே தவிர, முழுப்பொருளும் வெளிப்படையாகத் தெரிவதில்லை. பொருளை முழுமையாகப் புரிந்துகொள்ள, தொடர்களின் அமைப்பை முதலில் நமது மூளை தெரிந்துகொண்டு, அந்த அமைப்பின்மீது தனக்கே உரிய மொழித்திறன் மற்றும் உலகப் பொது அறிவு போன்றவற்றைச் செயல்படுத்துகிறது. மொழித்தொடர்களில் காணப்படுகிற சொல் அமைப்பு, தொடர் அமைப்பு போன்ற மொழி அமைப்புகள் (Language Structure) தொடர்கள் அடையாளப்படுத்துகிற பொருளுக்கேற்ப மாறுபட்டு அமைகின்றன. இந்த மொழி அமைப்புகளைப் பற்றிய மொழியறிவையும் தேவையான பிற உலகப் பின்புலப்பொதுஅறிவையும் கொண்டுதான் மனித மூளை, தொடர்கள் வெளிப்படுத்துகிற பொருளைப் புரிந்துகொள்கிறது. உலகப் பின்புலப் பொதுஅறிவை எவ்வாறு கணினிக்கு அளிப்பது என்பது வேறு ஒரு பிரச்சனை.

தொடர்நிலையில் பொருள் மயக்கம் (Syntactic Structural Ambiguity)

நாம் உருவாக்கும் மொழித்தொடர்கள் ஒவ்வொன்றும் ஒரே ஒரு அமைப்பைப் (Single Structure) பெற்றிருந்தால், அதை அடிப்படையாகக்கொண்டு, தொடர்ப்பொருளைப் புரிந்துகொள்வது சற்று எளிது. ஆனால் நடைமுறையில் இயற்கைமொழித் தொடர்கள் அவ்வாறு அமைவதில்லை. ஒரு தொடரில் ஒன்றுக்கு மேற்பட்ட அமைப்புகள் காணப்படலாம். அதாவது குறிப்பிட்ட தொடரில் இடம் பெறும் சொற்கள் தங்களுக்குள் வெவ்வேறு வகையில் இணையலாம். அப்போது பொருள் மயக்கம் ஏற்படுகிறது (Structural Ambiguity).

எடுத்துக்காட்டாக, “ நான் முரளியோடு கமலாவை இன்று பார்த்தேன்” என்ற தொடரை இரண்டுவகையில் புரிந்துகொள்ளலாம். “ நான் முரளியோடு “ “ கமலாவை இன்று பார்த்தேன்” என்றும் அல்லது “ நான்” “முரளியோடு கமலாவை இன்று பார்த்தேன்” என்றும் இரண்டுவகையில் புரிந்துகொள்ளலாம். ஒரே தொடரில் ஐந்து சொற்கள் தங்களுக்குள் எவ்வாறு இணைகிறது என்பதைப் பொறுத்துத்தான் தொடரின் பொருள் அமைகிறது. இதைத்தான் மொழியியலின் தொடரியல் ஆய்வு செய்கிறது. மேற்கூறிய தொடருக்கு

இரண்டு அமைப்புகள் இருந்தாலும், நாம் குறிப்பிட்ட சூழலை அடிப்படையாகக்கொண்டு, இரண்டு தொடர் அமைப்புகளில் ஒன்றைத் தேர்ந்தெடுத்து, பொருள் கொள்கிறோம். இதற்குத் தேவை, முதலில் குறிப்பிட்ட தொடரில் எத்தனை மற்றும் எந்தவகை அமைப்புகள் புதைந்துகிடக்கின்றன என்ற தொடரியல் அறிவும் (Syntactic Knowledge) , அதன்பின்னர் பின்புல அறிவை அடிப்படையாகக் கொண்டு இரண்டு அமைப்புகளில் ஒன்றைத் தேர்ந்தெடுக்கும் அறிவும் (Pragmatic knowledge) ஆகும்.

தொடர் நிலையில் (Syntax) மட்டுமல்லாமல், ஒலியனியியல் (Phonology), உருபனியியல் (Morphology) அல்லது சொல், பொருண்மையியல் (Semantics), கருத்தாடல் (Discourse) ஆகிய நிலைகளிலும் ஒரு தொடருக்கு ஒன்றுக்கு மேற்பட்ட அமைப்புகள் நிலவலாம். அதனால் ஒன்றுக்கு மேற்பட்ட பொருளுக்கு, அக்குறிப்பிட்ட தொடர் இடம் அளிக்கலாம். பொருள் மயக்கம் (Sense Ambiguity) ஏற்படலாம். இப்பொருள் மயக்கத்தைத் தீர்த்துவைத்தால்தான் (Disambiguation) , குறிப்பிட்ட தொடரின் குறிப்பிட்ட பொருளைக் கணினி புரிந்துகொள்ளமுடியும்.

உருபன் அமைப்புநிலையில் பொருள் மயக்கம் (Morphological Ambiguity)

“நான் கத்தி வரவழைத்தேன்” என்ற தொடரில் “ கத்தி” என்ற சொல்லானது “ கத்தி” என்ற ஒரு பெயர்ச்சொல்லைக் குறிக்கிறதா அல்லது “ கத்து + இ ” என்ற வினையெச்சத்தைக் குறிக்கிறதா என்பது பற்றி முடிவு எடுத்தால்தான், நாம் குறிப்பிட்ட தொடரின் பொருளைப் புரிந்துகொள்ளமுடியும். ஒரே சொல் இரண்டு அமைப்புகளுக்கு இடம் தருகிறது. “ வந்தவரை” என்பதின் அமைப்பு “ வந்த + வரை” என்பதா அல்லது “வந்தவர் + ஐ” என்பதா என்ற ஐயம் ஏற்படுகிறது. இது உருபனியல் நிலையிலான மயக்கம் எனப்படும்.

சொல் நிலையில் பொருள் மயக்கம் (Lexical Ambiguity)

“ பச்சைக் காய்கறி” , “ பச்சைப் பொய்” , “ பச்சை உடம்பு” ஆகிய தொடர்களில் “ பச்சை” என்ற சொல் மூன்று வேறுபட்ட பொருள்களைக் குறித்து நிற்கிறது. மொழி அகராதியில் “ பச்சை” என்ற சொல்லுக்கு இந்த மூன்று பொருள்களும் கொடுக்கப்பட்டிருக்கலாம். ஆனால் குறிப்பிட்ட தொடரில் மூன்றில் எந்தப் பொருளை எடுத்துக்கொள்வது என்பது அச்சொல்லுக்கு அடுத்துவருகிற சொல்லைப் பொறுத்து அமைகிறது. இது சொற்பொருண்மை நிலையிலான மயக்கம் எனப்படும்.

மேற்குறிப்பிட்ட பொருள்மயக்கங்களைத் தீர்த்துவைக்கக்கூடிய அறிவை - வழிமுறைகளை - எவ்வாறு கணினிக்கு அளிப்பது என்பது பற்றிய ஆய்வே கணினிமொழியியலாகும். சொல், தொடர் என்ற நிலைகளில் மட்டுமல்லாமல் ஒரு முழு உரை நிலையிலும் ஏற்படும் பொருள்மயக்கங்களை எல்லாம் தீர்க்கக்கூடிய மொழி அறிவைக் கணினிக்குக் கொடுக்கவேண்டும். அப்போதுதான் தமிழானது கணினித்தமிழாகப் பரிணமிக்கமுடியும்.

இலக்கணப் பொறியியலும் கணினி இலக்கணமும் (Grammar Engineering and Computational Grammar)

இதற்கு முதல் தேவை, தமிழ்மொழி அமைப்பை - பேச்சொலியில் இருந்து, முழு உரை வரை - முழுமையாக ஆய்வு செய்யவேண்டும். மனித மூளைக்கு எழுதப்படுகிற மொழி இலக்கணத்தைவிட அதிக விளக்கங்களை உடையதாக இது இருக்கும். ஏனெனில் மனித மூளைக்கு உள்ள உலகப் பின்புல அறிவு கணினிக்குத் தற்போது இல்லை. இதுபோன்ற ஒரு மொழி இலக்கணத்தை உருவாக்கும் பணியானது (மொழி) இலக்கணப்பொறியியல் என்று அழைக்கப்படுகிறது. இந்த அடிப்படையில் இன்றைய தமிழுக்கு ஒரு இலக்கணம் - கணினி இலக்கணம் உருவாக்கப்படவேண்டும். நாம் வெளிப்படுத்த விரும்பும் கருத்தை அல்லது பொருளைத் தமிழ்மொழியானது தனது ஒலியனியல், உருபனியல், தொடரியல், பொருண்மையியல், பின்புல அறிவியல், கருத்தாடல் இயல் போன்ற பல நிலைகளில் எவ்வாறு அடையாளப்படுத்தி நிற்கிறது என்ற மொழி அறிவை இந்தக் கணினி இலக்கணம் வெளிப்படுத்தி நிற்கும்.

மேற்குறிப்பிட்ட இலக்கணம் கணினிக்கான இலக்கணம் என்பதால், ஒரு மின்னணுக் கருவியான கணினிக்குப் புரியக்கூடிய வகையில் அது அமையவேண்டும். கணித அடிப்படையில் (Mathematical) அமையவேண்டும். தமிழ் மொழி அமைப்பை - பேச்சொலியிலிருந்து உரை அமைப்பு வரை - வெளிப்படுத்தும் முறைசார் இலக்கணமாக (Formal Grammar) அமையவேண்டும். இதற்கான பல நுட்பங்களைத் தற்கால மொழியியல் துறையில் இருந்து மட்டுமல்லாமல், கணிதவியல், இயற்பியல், புள்ளியியல், கணினியியல், மின்னணுவியல் போன்ற துறைகள் அளிக்கிற அறிவிலிருந்தும் பெறவேண்டியுள்ளது.

கணினிமொழியியல் நோக்கில் தமிழ் ஆராய்ச்சிப் பணி (Research in Tamil Computational Linguistics)

பல துறைசார்ந்த ஒரு துறையாக இன்று கணினிமொழியியல் துறை வளர்ந்துவருகிறது. இதன் அடிப்படையில் தமிழ்மொழி அமைப்பை - அறிவை- கணினிக்கு அளிப்பதன்மூலம்தான், கணினித்தமிழ் வளரமுடியும். தமிழ்மொழி அறிவைக் (Knowledge of Tamil Language) கணினிக்கேற்ற வகையில் தரவு அமைப்பு மூலமும் கணினிநிரல்வழிமுறை மூலமும் கொடுக்கும் பணியே கணினித்தமிழ் ஆராய்ச்சிப்பணியாகும். இதில் மிக அடிப்படையானவை இரண்டு : ஒன்று, தமிழ்மொழி இலக்கணம் (Grammar), மற்றொன்று பகுப்பி (Parsers). ஒரு தமிழ் உரையைத் தமிழ் இலக்கணத்தின் அடிப்படையில் பகுப்பிகளைக் கொண்டு பகுப்பாய்வு செய்யும்போதுதான் , அந்த உரையில் புதைந்திருக்கும் அமைப்பைக் கண்டறிந்து, அதற்கேற்ற பொருள் விளக்கத்தைக் கணினியால் பெறமுடியும்.

இப்பணிக்குத் தேவையான துறைசார் அறிவைக் கணினிமொழியியல் இன்று அளித்துள்ளது. பேச்சொலிகள் முதல் ஒரு முழுமையான கருத்தாடல்வரை ஒரு மொழியைக் கணினிநோக்கில் எவ்வாறு ஆராய்வது என்பதற்கான பல வழிமுறைகளை அது உருவாக்கிவருகிறது. பேச்சொலி உருவாக்கி (Speech Synthesizer) , பேச்சொலி ஆய்வாளர் (Speech Analyzer) சொல் பகுப்பி (Morphological Parser), தொடர்பகுப்பி , இலக்கண வகைப்பாட்டுக் குறியீட்டாளர் (Word Class Tagger), பொருண்மை ஆய்வாளர் (Semantic Analyzer), (Syntactic Parser) பொருள்மயக்கத் தெளிவாக்கி (Word Sense Disambiguator) போன்ற பல்வேறு மொழி மற்றும் பேச்சு ஆய்வுக் கருவிகள் உருவாக்கப்பட்டுள்ளன. மொழித்தரவுகளை (Language Corpus) உருவாக்கத் தேவையான அடிப்படையான வழிமுறைகள் தரவுமொழியியலில் (Corpus Linguistics) முன்வைக்கப்பட்டுள்ளன. கணினிக்கேற்ற மொழி அகராதிகளை (Computational Lexicon) உருவாக்குவதற்குத் தேவையான பல வழிமுறைகள் - சொல்வலை (WordNet), உருவாக்க அகராதி (Generative Lexicon) போன்றவை முன்வைக்கப்பட்டுள்ளன.

கணினிமொழியியலின் வளர்ச்சி

உலக அளவில் கணினிமொழியியல் துறையானது இன்று நன்கு வளர்ந்துள்ளது. பிற துறை அறிவுகளையும் அடிப்படையாகக்கொண்டு, பல்வேறு ஆய்வுக்கோட்பாடுகளையும், மாதிரிவடிவங்களையும் அது முன்வைத்துள்ளது. ரெகுலர் எக்ஸ்பிரஷன் (Regular Expression) , ஃபைனைட் ஸ்டேட் ஆட்டோமேட்டா (Finite-State Automata) , இயந்திரக் கற்றல் (Machine Learning) , ஹிட்டன் மார்க்கோவ் மாடல் (Hidden Markov Model), என்-கிராம் வழிமுறை (N-gram Method) எனப் பல கணினிய வழிமுறைகளைக்கொண்டு, இயற்கைமொழித்தொடர்களை மின்னணு உரைவடிவத்தில் வெளிப்படுத்தி ஆய்வு செய்யும் கருவிகளை அது உருவாக்கியுள்ளது.

மொழியியல் துறையில் உருவாக்கப்பட்டுள்ள இருநிலை உருபனியியல் (Two-Level Morphology), மாற்றிலக்கணம் (Generative Grammar), பொதுத் தொடர் அமைப்பு இலக்கணம் (Generalised Phrase Structure Grammar - GPSG), சொல்-செயல்பாட்டு இலக்கணம் (Lexical - Functional Grammar - LFG), கிளை-ஒட்டு இலக்கணம் (Tree -Adjoining Grammar - TAG), தலைமை சார் தொடரமைப்பு இலக்கணம் (Head-Driven Phrase Structure Grammar - HPSG) போன்ற பல்வேறு இயற்கைமொழி இலக்கண

மாதிரிவடிவங்களும் கணினிமொழியியலில் மிகவும் பயன்படுகின்றன. மேலும் புள்ளியியல் அடிப்படையிலான கணினிமொழியியல் ஆய்வும் (Statistical Computational Linguistics) , குறிப்பாக நிகழ்ச்சித்தகவு புள்ளியியல் (Probability Statistics) அடிப்படையிலான கணினிமொழியியல் பெருமளவில் பயன்படுகிறது.

கணினிமொழியியலின் அடிப்படையில் தமிழ் எழுத்து வழக்கு, பேச்சு வழக்கு இரண்டையும் ஆய்ந்து, ஒரு சிறந்த கணினித்தமிழ் இலக்கணத்தை - தமிழ்த்தரவு , பேச்சொலி உருவாக்கி, பேச்சொலி ஆய்வாளர், சொல் மற்றும் தொடர் பகுப்பிகள், கணினி அகராதி உட்பட - நாம் உருவாக்கினால்தான், இன்றைய மொழித்தொழில்நுட்பத்தின் பயன்களை நாம் பெறமுடியும். சொல்திருத்தியிலிருந்து தானியங்கு மொழிபெயர்ப்பு வரை பல்வேறு பயன்பாட்டு மொழிக்கருவிகளைத் தமிழுக்கு உருவாக்கமுடியும். இதற்கான சில அடிப்படைப் பணிகள் ஏற்கனவே தமிழகத்தில் மேற்கொள்ளப்பட்டுள்ளன.

உயர்நிலைக் கல்வி மற்றும் ஆய்வு நிலையில் கணினிமொழியியல் மற்றும் மொழித்தொழில்நுட்பம் தமிழகத்தின் பல்வேறு பல்கலைக்கழகங்களிலும் கல்லூரிகளிலும் அறிமுகப்படுத்தப்படவேண்டும். குறிப்பாக, உயர்நிலைத் தமிழ்மொழிக்கல்வியில் கணினிமொழியியல் பாடம் வைக்கப்படவேண்டும். கணினிமொழியியலில் மனிதவளம் பெருமளவில் உருவாக்கப்படவேண்டும். இதுபோன்ற திட்டமிட்ட கணினித்தமிழ் வளர்ச்சித்திட்டத்தின்மூலமே செம்மொழித்தமிழானது இன்றைய அறிவியல் உலகிற்கேற்ற கணினித்தமிழாகவும் வளர்ந்து, உலக அளவில் தலைசிறந்து விளங்கும்.

தேர்ந்தெடுக்கப்பட்ட துணைநூற்பட்டியல்

1. Allen, James (2003) **Natural Language Understanding** , Pearson Education: India
2. Cole, Ronald et al (ed.) (1997) **Survey of the State of the Art in Human Language Technology** , CUP : New York
3. Dale, Robert., H. Moise & Harold Somers (ed.) (2000) **Handbook of Natural Language Processing** , Marcel Dekker: New York
4. Grishman, Ralph (1999) **Computational Linguistics - An Introduction** , CUP : New York
5. Housser, Roland (2001) **Foundations of Computational Linguistics**, Springer: London
6. ----- (2006) **A Computational Model of Natural Language Communication**, Springer: London
7. Jurafsky, Daniel & James H. Martin (2003) **Speech and Language Processing**, Pearson Education : India
8. Manning, C . & H.Schutze (1999) **Foundations of Statistical Natural Language ,Processing** , MIT: Cambridge
9. Manoharan, M. (2009) **Globalization and Language Planning (With Special Reference to Modern Tamil)** , Unpublished Ph.D. dissertation , University of Madras: Chennai
10. Mitkov, Ruslan et al (2003) **The Oxford Handbook of Computational Linguistics** , OUP: London

Noun Phrase Chunker using Finite State Automata for an Agglutinative Language

Vijay Sundar Ram R and Sobha Lalitha Devi

AU-KBC Research Centre,
MIT Campus of Anna University,
Chromepet, Chennai -44
{sobha,sundar}@au-kbc.org

Abstract

This paper presents a system for noun phrase chunking for Tamil, an agglutinative language. The partial chunking of the text is done by a rule-based approach where the rules are embedded in a finite state automaton (FSA), which recognizes the chunks at a high accuracy rate and speed. The chunking of text being the pre-processing task needs to be of good performance. The evaluation of the system shows a recall of 93.7% and precision of 94.9%.

Introduction

We present a system for noun phrase chunking for an agglutinative language, Tamil. The noun phrase considered here has a head noun preceded by determiner, quantifier, classifier and adjective in sequence. Here the determiner, quantifier, classifier, and adjective are optional. The recursive noun phrase occurs when possessive case marker occurs. Often the noun phrase has the information about the happening. The event executor and event receiver of the event are usually noun phrases around the verb. The proper chunking of noun phrase helps in improving the efficiency of the information extraction, machine translation and in information retrieval system by improving the terms in the term vs document matrix. The system is built using rule based approach, where the rules are embedded in a finite state automaton (FSA), as the structures are recognized with high degree of accuracy.

Several methods have come up for this task. Church's stochastic noun phrase tagging was one of the early attempts, where the corpus frequencies were used to determine the noun phrase boundaries (Church, 1988). Abney did partial parsing using finite state cascades, where the finite state cascade has sequence of levels. Phrase at one level is built on the phrase at the previous level without any recursion (Abney, 1996). Ramshaw and Marcus used Eric Brill's transformation based learning for recognizing the noun chunks and other text chunking (Ramshaw, 1995). Dimitrios Kokkinakis and Sofie Johansson Kokkinakis did a cascaded Finite-State Parser for Syntactic Analysis of Swedish (Dimitrios, 1999). Chunk tagger using markov model technique for recognizing the internal structures and syntactic category of simple as well as complex structures was done by Wojciech (Wojciech 1998). Noun phrase chunking for German using probabilistic context-free parser for learning and tagging the most probable chunk sequence is done by Helmut (Helmut, 2000).

The rest of the paper is organised as follows. The section 2 starts with briefing about the agglutinative language with samples of noun phrases. In section 3 techniques of finite state automata (FSA), different preprocessing works and the implementation of the Tamil NP chunker are explained. The evaluation and the discussions on result in section 4 and finally the paper ends with conclusion.

Description about the Language

The system is developed for an agglutinative language such as Tamil, Telugu, Malayalam, etc. Here the language under consideration is Tamil. Tamil belongs to the South Dravidian family of languages. It is a verb final language and allows scrambling. It has post-positions, the genitive precedes the head noun in the genitive phrase and the complementizer follows the embedded clause. Adjective, participial adjectives and free relatives precede the head noun. It is a nominative-accusative language like the other Dravidian languages. The subject of a Tamil sentence is mostly nominative, although there are constructions with certain verbs that require dative subjects. Tamil has png agreement.

Noun Phrases in Tamil

Tamil is a relatively free word order language, but when it comes to noun phrases and clausal constructions it behave as a fixed word order language. As in other languages, Tamil also has optional and obligatory parts in the noun phrase. Head noun is obligatory and all other constituents that precede the head noun are optional. In this section we discuss in detail about different noun phrases in Tamil. Consider the following

1. periya viitu
big+ADJ house+N+NOM
(Big House)
2. oru azhakiya viitukku
one +Q beautiful+ADJ house+N+DAT
(For a beautiful house)
3. inta koyil
this+DET temple+N+NOM
(This temple)

In examples 1, 2 and 3 the structure of the noun phrase has a head noun, which may or may not be preceded by optional categories like determiner, quantifier, classifier, and adjective. This is the general structure of noun phrase. This structure is represented by the following Rule.

Rule a:

NP=>[Determiner][Quantifier][classifier] [Adjective] {N}

- 4.avanutaiya bramaandamaana araiyin vizakukal His+PN+POS grand+ADJ
room+N+POS lamp+N+PL
(His grand room's lamp)

Example 4 has a possessive noun. If the head noun has a possessive case marker and followed by a noun phrase, both the noun phrases are chunked into one noun phrase. This may go recursively. This is shown in Rule b.

Rule b:

possessive NP => [Determiner][Quantifier][classifier] [Adjective] {N+possessive case}

NP=> (possessiveNP)* Rule a

The rule a and b are combined into a single rule

Rule 1:

NP=> [possessive NP] +Rule a

5. ivviitu
this+DET house+N+NOM
(This house)

Here compounding of the determiner and the noun forms single word. The tags of this word also obey the previous rule but the tags are in the same word. This type is handled separately to accept the tags within the word.

6. powuc ceyalaalar
(Secretary)

In example 6, here the two nouns, which are suppose to be a single noun is written separately and connected by a sandhi. This is handled by the rule given below:

Rule 2:

NP=> {N+sandhi} {head N}

7. therku maaligai
south+N+NOM palace+N+NOM
(Southern palace)

Example 7 has two nominative nouns, where the first denotes the direction. These two nouns go together to form a noun phrase. The eight directions north, south, east, west, northeastern, northwestern, southeastern and southwestern are considered. This is handled by the following rule

Rule 3:

NP => {direction denoting NP} {NP}

8. kadal niir
sea+N+NOM water+N+NOM
(Seawater)

When two nominative nouns are adjacent to each other the two nominative nouns can be chunked into noun phrase.

NP => {N+NOM} {N+NOM}

Tamil NP Chunker

The noun phrase chunking is done by building finite state automata using the linguistic rules. By traversing the text through the FSA, the chunking of the text is done.

Finite State Automata

A finite state automaton is an abstract device used for recognizing simple syntactic structures or patterns in text strings. An automaton is normally depicted by a directed graph in a so-called state diagram. An FSA as a string-processing device accepts text strings as input and decides if the structure is correct, that is, it either accepts or rejects the string. From a mathematical point of view, it may therefore be regarded as a function, mapping a set of strings to the set {ACCEPT, REJECT}. Based on the transitions given by the FSA, they are classified as Non-Deterministic Finite State Automata (NFA) and Deterministic Finite State Automata (DFA).

The deterministic FSA is found to be unsuitable for parsing of the sentence. Since the grammar of the languages are context free and non-deterministic in nature. But for partial chunking such as noun phrase chunking the structure of the phrase are highly fixed. So usage of deterministic FSA helps in high degree of accurate chunking.

The Deterministic Finite-State Automaton (DFA) which is a special case of NFA has the following requirements:

1. There are no transitions involving ϵ .
2. No state has two outgoing transitions based on the same symbol

The Deterministic FSA used in this task is shown in Figure 1.

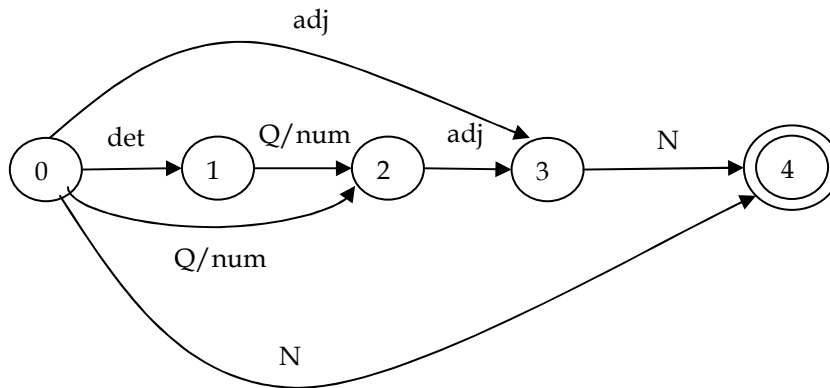


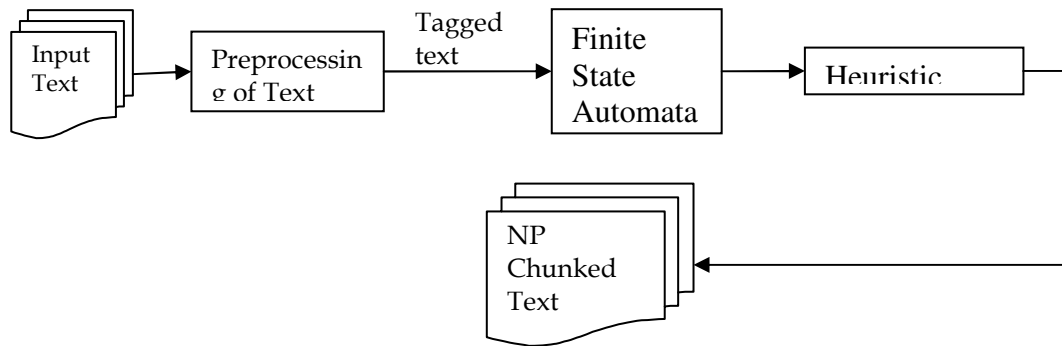
Figure 1 Deterministic FSA

In the present work the FSA is used in the form of a state table. The morphological tags of the words are the transition symbols.

Architecture of the System

The architecture of the system is shown in figure 3.

Figure 3 Architecture



The Preprocessing Works

The input text to be chunked is initially fed to the morphological analyser (Viswanathan 2003), which gives multiple outputs for each word. The morphologically tagged output is sent to a hybrid approach based Part of Speech (POS) tagger (Arulmozhi 2006), where the ambiguities in the output of the morphological analyser are corrected using the POS tags.

On giving the preprocessed text, which is now a tagged text, state transition takes place in FSA based on the tags as the symbol to trigger the transition. The chunking of the phrase completely depends on the morphological tags.

The process in Finite State Automata

The FSA is built using the manually crafted linguistic rules.

When a tagged sentence is given, the current word's tag is considered as the transition symbol to have transition to the next state, in the next state the next word's tags is considered. Similarly the traversing in FSA happens till it reaches the end state. If it successfully reaches the end state this part of the text is chunked as noun phrase.

The nouns with all the case markers have the same noun phrase structure except the nouns with possessive case marker. So the nouns with all case markers except the possessive case marker are treated similarly. So the case marking tags in the nouns except the possessive case are not considered. This suppression of the sub-tags reduces the number of transition from 400 to 40 in the determinized FSA. This optimizes the FSA and improves the speed and the efficiency of the chunking process.

The noun phrase chunked text obtained after traversing through the FSA, is again processed using the heuristic rules.

Results and Discussion

The system is evaluated with the data taken from the CIIL corpus (Central Institute of Indian languages corpus). The numbers of sentences under consideration are 500, which contains 2180 noun phrases. The

number of noun phrases recognized correctly by the system is 2043, which is at a precision of 94.9% and the recall of 93.7%. The results are tabulated as follows.

Table 1 Performance of the Chunker

S.No	Number of Sentences	Number of NPs present	Number of NPs chunked	Number of NPs correctly chunked	Recall %	Precession %
1	500	2180	2153	2043	93.7	94.9

The capitalization of the first letter gives an easy identification of proper nouns in English language. This is not there in Indian languages, which makes the task of recognizing the noun phrase complex. The challenging task is chunking the noun phrase formed by two nouns such as “kadal niir”(sea water), which should be recognized as one chunk. We cannot go blindly by chunking two nominative nouns into a chunk because doing that will make error in the following case.

raman palam saappittaan
 raman+N+NOM fruit+N+NOM eat+V+PST
 (raman ate fruit)

Here if we apply the rule that two adjacent nominative nouns will form a chunk, then the two nouns ‘raman’ and ‘palam’ will form as one chunk, “raman palam”. But this is an incorrect chunk since they are two different noun phrases where ‘raman’ is the subject and ‘palam’ is the object of the sentence. The accusative marker, the marker of direct object, in ‘palam’ is dropped which is a common phenomenon in Tamil. We are trying to over come this by subcategorizing the verb to identify the subject and object of the verb. These types of correction have to be done after the chunking is over.

Conclusion

The paper presents a noun phrase chunking for an agglutinative language, Tamil. In developing the system we are using finite state automata technique, where the rules are embedded in it. The system performs with 94.9% precession and 93.7% recall. The system takes 2.5 sec to chunk 1000 sentences. In future, we will be handling the noun phrases with multiple nouns.

Reference

1. Steven Abney (1996). Partial Parsing via Finite-State Cascades. In *Proceedings of the ESSLLI '96 Robust Parsing Workshop*
2. Arulmozhi Palanisamy and Sobha Lalitha Devi. 2006. *HMM based POS Tagger for a Relatively Free Word Order Language*, Journal of Research on Computing Science, Mexico. 18:37-48
3. Dimitrios Kokkinakis and Sofie Johansson Kokkinakis 1999. in *Proceedings of EACL'99*, Bergen
4. Kenneth W. Church. 1988. A Stochastic Parts Program and Noun Phrase Parser for unrestricted Text. In *Proceedings of the Second Conference on Applied Natural language Processing*

5. L. Ramshaw and M. Marcus. 1995. Text Chunking using Transformation-Based Learning. In *Proceedings of the Third Workshop on Very Large Corpora*
6. Helmut Schmid and Sabine Schulte im Walde 2000. Robust German Noun Chunking With a Probabilistic Context-Free Grammar. In *proceedings of COLING 2000*, Saarbrücken
7. Wojciech Skut, Thorsten Brants (1998) Chunk Tagger Statistical Recognition of Noun Phrases. In the *proceedings of ESLLI-1998 Workshop on Automated Acquisition of Syntax and Parsing*. Saarbrücken, Germany
8. Viswanathan, S Ramesh Kumar, B Kumara Shanmugam, S Arulmozi 2003, A Tamil Morphological Analyser , *International Conference On Natural Languageprocessing*, Central Institute of Indian Languages, Mysore
9. Dr. Steve Sugden <http://www.it.bond.edu.au/inft150/033/lectures/ch10.pdf>

Animated Sangathamizh Poems - E Learning

Arul Natarajan
dr.n.arul@gmail.com

Introduction

This paper on **Animated Sangathamizh poems** is an attempt to

- Understand the importance of Sangathamizh poems
- Revive the vibrant and dynamic values of Tamil culture and heritage
- Introduce it to the youth of modern world
- With the latest technology available at our hands.

Tamil

Tamil is a language with a literature that is classical— i.e., it is ancient, it has an independent tradition that arose mostly on its own, not as an offshoot of another tradition, and has a large and extremely rich body of ancient literature.

The quality of classical Tamil literature is such that it is fit to stand ahead of the great literatures of Sanskrit, Greek, Latin, Chinese, Persian and Arabic. The subtlety and profundity of its works, their varied scope, and their universality qualify Tamil to stand as one of the great classical traditions and literatures of the world. Tamil is one of the primary independent sources of modern Indian culture and tradition.

Sangathamizh Literature

Sangam literature refers to classical Tamil literature created between the years c. 600 BCE to 300 CE. This collection contains 2381 poems composed by 473 poets, some 102 of whom remain anonymous. Sangam literature is primarily secular dealing with everyday themes in a Tamil context. Much of the Tamil literature believed to have been composed in the Sangam period is lost to us, though detailed lists of works known to the 10th century compilers have survived.

The poems belonging to the Sangam literature were composed by Tamil poets, both men and women, from various professions and classes of society. These poems were later collected into various anthologies, edited, and with colophons added by anthologists and annotators around 1000 CE.

Sangam Poems falls into two categories: the 'inner field' (Agam), and the 'outer field'(Puram) as described even in the first available Tamil grammar, the *Tolkappiyam*.

The 'inner field' topics refer to personal or human aspects, such as love and sexual relationships, and are dealt with in a metaphorical and abstract manner.

The 'outer field' topics discuss all other aspects of human experience such as heroism, valour, ethics, benevolence, philanthropy, social life, and customs.

Sangathamizh illustrates the thematic classification scheme first described in the *Tolkappiyam*

The classification ties the emotions involved in agam poetry to a specific landscape. These landscapes are called *thinai*. These are: *kurinji* mountainous regions; *mullai* forests; *marutham* agricultural land; *neithal* coastal regions; *paalai* deserts. In addition to the landscape based *thinais*, *kaikkiLai* and *perunthinai* are used for unsolicited love and unsuited love respectively.

Similar *thinais* pertain to *puram* poems as well, though these categories are based on activity rather than landscape: *vetchi*, 'karanthai, vanchi, kanchi, umignai, nochchi, thumbai, 'vaagai, *paataan*, and *pothuviyal*.

Revival

Tamil, being one of the most ancient languages, yet vibrant and dynamic with its inexhaustible thoughts in every realm influencing humanity through its arts and literature has provided guidelines for happy living, just governance, noble behavior etc. It's our duty to take this torch further in our time for the future by all the modern means.

The works of Sangam literature were lost and forgotten for several centuries. A revival took place from the late nineteenth century when works of religious and philosophical nature were written in a style that made it easier for the common people to enjoy. Several Tamil scholars such as S. V. Damodaram Pillai and U. V. Swaminatha Iyer. They painstakingly collected and catalogued numerous manuscripts in various stages of deterioration. They printed and published Tholkappiyam, Nachinarkiniyar urai (1895), Tholkappiyam Senavariyar urai, (1868), Manimekalai (1898), Cilappatikaram (1889), Pattupattu (1889), and Purananuru (1894), all with scholarly commentaries. Damodaram Pillai and Swaminatha Iyer published more 100 works in all, including minor poems. Nationalist poets like Bharathiyar and Bharathidasan began to utilize the power of poetry in influencing the masses.

With growth of literacy, Tamil prose began to blossom and mature. Short stories and novels began to appear. Many scholars including our Hon'ble CM have written books on sangathamizh poems. The popularity of Tamil Cinema has also provided opportunities for modern Tamil creative heads to take up sangathamizh in. Movies like Thiruvilayadal, Mannadhimannan, Iruvar and so on, has references to sangam poems.

Why Sangathamizh Poems?

Sangam Poems are the authority for moral Life. They fall into two categories; Agam and Puram, referring to personal or human aspects, such as love and sexual relationships in the former and all other aspects of human experience such as heroism, valor, ethics, benevolence, philanthropy, social life, and customs in the latter. This cultural treasure is known only to Tamil scholars and is still out of reach of common mass. This is because of lack of awareness of our heritage and culture and above all poor education. Though some of these poems are taught in schools and colleges, there is still no proper way to inspire the learner to have a further reading of the literatures.

Animated Sangathamizh Poems

Animated Sangathamizh Poems would be the adaptation of classical Tamil poems to modern music and animations without losing the values and traditional tastes. Poems from Ettuthogai, Pathu Paattu, Padhinenkeezhkanakku etc., are full of interesting scenes paving way for good visualisation of the glorious living of our Tamils. These poems can be visualised in animated format which will attract not only every Tamil, but also every human being in the globe who try to attain the ethical values.

How to implement

Since the need of understanding cultural values in the society are increasing day by day, Animated Sangathamizh poems with musical rendering should reach the public through all the technological means available update. The challenge is quality, process and technology we are going to use so that it reaches the mass with a greater impact. The media in hand are Satellite channels, CDs, iphones and mobile phones. Since technology is improving day by day, High quality music, Stunning graphics and animations are possible to achieve the objective.

Music: Differing from the traditional way of composing music, the Animated Sangathamizh poems will have tunes to the taste of youth to attract them, at the same time without compromise in the relevance and values. Various music composers have proved that this is possible.

Animation: Stunning animations can be made for sangathamizh poems as they naturally have wonderful visuals and descriptive storyboards of past glory. Care will be taken to make the animations and graphic interesting, relevant to present day situations.

Meaning: The ultimate objective of the attempt is to reach this to the common mass. This can be achieved by using simple language for describing the meaning of the poems. People need not be literate to understand what is being said.

Advantages

- **A great revolution:** This will be indeed a great revolution in revival of Sangam Literatures because through this attempt we can uplift the values of life which is found losing importance in the modern, fast world.
- **Multimedia:** Throughout history, man has shown his fascination towards visualizing his creative thoughts in various available resources. The technologies available with us can attract audience of all stages of the social pyramid.
- **Teaching/Learning aid:** This can further be an excellent resource for researchers, teachers and learners where hours of time spent on by hearing text form can be reduced to a minimal time with permanent registry in mind.
- **Future Generations:** This will certainly be a treasure for the generations to come as the world is becoming more and more modern and there is no time to turn back to even glance at the values which are missing.

Conclusion

In spite of being spoken by a whole race of people including the illiterate among us, Tamil has maintained such a continuity and uniformity that literature written 2000 years ago can be understood by the educated readers of the present. But we should not stop with this. Every common man, literate or illiterate should enjoy the essence of Tamil and we believe that Sangathamizh poems - being the most ancient literature, yet with inexhaustible thoughts in every realm influencing humanity through its words has provided guidelines for happy living, just governance, noble behaviour etc. It's our duty to take this further in our time for the future and one such way is Animated Sangathamizh Poems.

Representation of Kinship in WordNet

S.Arulmozi

Dravidian University

Abstract

WordNet (Fellbaum,1998) is one of the most resourceful semantic lexicons in English. Its main advantage is that it is hand-crafted, so data stored within its semantic network are of high quality. It is used in most of the NLP applications, particularly in sense disambiguation tasks. WordNets are already available in most of the languages of the world including Hindi. Efforts are underway in Indian languages using the expansion approach with Hindi WordNet as the base. Kinship presents a tough problem in the construction of WordNet, especially in Tamil and other major Dravidian languages. This paper presents the lexicographical issues involved in the construction of synsets (synonym sets) in general and kinship hierarchy in particular. A brief account on the representation of kinship hierarchy in WordNet will also be provided

Introduction

The Princeton English WordNet (Fellbaum, 1998) is one of the most resourceful semantic lexical database in English. Its main advantage is that it is hand-crafted, so data stored within its semantic network are of high quality. It is widely used as a resource in many NLP applications such as Information Retrieval, Word Sense Disambiguation, etc.

The continuous expansion of the multilingual information society with a growing number of new languages present on the Web has led in recent years to a pressing demand for multilingual applications. To support such applications, multilingual language resources are needed, which however require a lot of human effort to be built. For this reason, the development of language independent resources which factorize what is common to many languages, and are possibly linked to the language-specific resources, could bring great advantages to the development of the multilingual resources in Indian languages.

Princeton's English WordNet inspired extensive development of WordNets in European languages, EuroWordNet (Vossen,) and also in other languages across the globe including WordNets in Indian languages, IndoWordNet (Pushpak,).

In this paper, a brief account on WordNet and construction of synsets is given. The paper is organized as follows: Section 2 details about WordNet and activities in Indian languages. Section 3 deals with the construction of synsets in general and pinpoints few problems faced during the construction of synsets. Section 4 briefly lists the ontology of kinship in English WordNet followed by the problems faced in creating synsets for kinship concepts in Tamil and Telugu. The last section summarizes the work.

WordNet

WordNet was originally conceived and developed as a lexical database for English on the basis of psycholinguistic properties. The major lexical categories such as nouns, verbs, adjectives and adverbs are organized in terms of sets of synonyms (synsets) each representing a lexical concept.

A synset is a set of synonyms (word forms that have the same or similar meaning) and two words are said to be synonymous if their mutual substitution does not alter the truth-value of a given sentence in which they occur, in a given context. For example, {computer, computing machine, computing device, data processor, electronic computer, information processing system} form a synset because they can be used to refer to the same concept. These synsets are interconnected by certain relations, lexical relations such as synonymy, antonymy and semantic relations such as hyponymy (between specific and more general concepts) and meronymy (between parts and wholes).

An example of a synset is reproduced (from WordNet 2.1) here for clarity:

The synset for {computer, computing machine, computing device, data processor, electronic computer, information processing system} is related to:

- more general concept or the hypernym synset {machine}
- more specific concepts or hyponym synsets {analog computer}; {digital computer}; {node, client, guest}; {number cruncher}; {pari-mutuel machine, totalizer, totaliser, totalizator, totalisator} and {server, host}
- parts it is composed of {busbar, bus}; {cathode ray tube, CRT}; {central processing unit, CPU, processor, mainframe};

Each of these synsets is again related to other synsets as is illustrated for {machine} that is related to {device}, and {CPU} that is related to other parts {mother board, CPU board}, {circuit, electrical circuit}.

WordNet (2.1 version) has approximately 120,000 lexical items (word forms) organized into 100,000 meanings (word meanings). For most of the synset, a brief definition (gloss) is provided.

The success of the English WordNet has paved way for the emergence of several projects with the aim constructing WordNets in various languages and developing multilingual WordNets. EuroWordNet, a conglomeration of WordNets in European languages is an important project that has come up with a multilingual WordNets. Similar efforts are underway in Indian languages. Hindi WordNet is leading the way for all Indian language WordNets under IndoWordNet.

WordNet building activities in Dravidian languages started with the work of Tamil WordNet²⁸ at AU-KBC Research Centre using Rajendran's (2001) ontological classification of Tamil vocabulary. Work on Dravidian WordNet (comprising WordNets in four major Dravidian languages, viz. Kannada, Malayalam, Tamil and Telugu) started during a Workshop²⁹ held at Chennai in which synsets were built for Construction Domain. Currently Dravidian WordNet³⁰ activity is being carried out for Kannada at University of Hyderabad, Malayalam at Amrita Vishwa Vidyapeetham, Tamil at Tamil University and Telugu at Dravidian University.

²⁸ Project partially funded by Tamil Virtual University.

²⁹ Workshop on WordNet for Dravidian Languages organized from 2-3 June 2003

³⁰ Project funded by the Ministry of HRD, Govt. of India.

Construction of Synsets

Various approaches are followed in the construction of WordNets across the languages of the world. For the Indian languages, WordNets are constructed using the expansion approach. For the construction of WordNet in Tamil and Telugu, we also follow the expansion approach, i.e. Hindi WordNet synsets are taken as a starting point of departure. The concepts provided along with the Hindi synsets are first conceived and appropriate concepts in Telugu are manually provided by language experts. The Telugu synsets are then built based on the concepts created keeping in view the three principles, viz. Minimality, Coverage and Replaceability.

Below we present the challenges/problems faced in the construction of core synsets in general and kinship synsets in particular.

Lexicographical Concerns

As mentioned earlier, Hindi concepts are first conceived and appropriate concepts are provided along with synsets in target languages. But during this process, problems occurred when we are faced with concepts that have no appropriate equivalents. For example,

Concept (HWN ID 7531): चालीस सेर की एक तौल

cAlIsa sera kI eka taula `a measure of 40 kg'.

For this concept, there is no corresponding equivalent in Tamil or Telugu. But there are varying usages in different dialects. For instance, in Kuppam, the measure is equal to 10 kg whereas in Kadapa district of Andhra Pradesh it is 14 kg, whereas in Tamil Nadu it is 10kg. When it comes to providing equivalent synsets, Hindi and Telugu and Tamil uses the same, i.e. *maNu*.

Concept (HWN ID 24): मादा शेर

mAdA Sera `a female tiger'

For the above concept, there is no problem is assigning equivalent concepts and synsets, but when it comes to providing equivalent sentences (which we mostly translate for developing parallel corpora), we come across difficulty. In most cases, we do not use gender while providing translations. This is not the case with this concept alone, but in all the concepts involving gender. For example, female rat (HWN ID 335), female parrot (HWN ID 1278), etc.

Kinship in WordNet

In English WordNet, kinship is represented in the following way: Kinship is a kind of relatedness or connection by blood or marriage or adoption. It can be classified as: 1) **affinity or phylogenetic relation** which is in biology means a state of relationship between organisms or groups of organisms resulting in resemblance in structure or structural parts; 2) **descent, line of descent, lineage, filiation** which is the kinship relation between an individual and the individual's progenitors; 3) **affinity** which in anthropology is the kinship by marriage or adoption and not a blood relationship; 4) **consanguinity or blood kinship** which in anthropology is the kinship that is related by blood; 5) **parentage or birth** which is the kinship relation of an offspring to the parents; 6) **fatherhood** which is the kinship relation between

an offspring and the father; 7) **motherhood** which is the kinship relation between an offspring and the mother; 8) **sisterhood** which is the kinship relation between a female offspring and the siblings; 9) **brotherhood** which is the kinship relation between a male offspring and the siblings; and 10) **marital relationship** which is the relationship between wife and husband.

When we come to WordNet building activities in Indian languages, Hindi WordNet which is the pioneer is taken as the source language for building WordNets. That is, Hindi WordNet's concepts are taken as a starting point and WordNets are getting built using the expansion approach. In the outset, this approach looks trivial and economical considering the interlinking of synsets of different languages. But when it comes to kinship relations, however, Hindi concepts create problems while assigning synsets in Dravidian languages, especially Tamil and Telugu. Let us examine a few in the following pages.

Problems in creating synsets involving kinship relations

In all, there are 54 concepts in Hindi WordNet which involve kinship relations. For the purpose of this paper, only the problematic concepts are taken into consideration which needs special attention.

1. HWN ID 7379: भाई का लड़का /bhai ka ladka/ `brother's son'

Synsets: "भतीजा, भ्रातृज, भ्रातापुत्र, भ्रातृपुत्र, भतीज, अवतंस, अवतन्स"

/BatIjA, BrAtRuja, BrAtAputra, BrAtRuputra, BatIja, avataMsa, avatansa/

When it comes to Tamil, providing concept is a problem. Straightforward, one can assign சகோதரன் மகன். But when assigning synsets, one comes across ambiguity in the concept. i.e. which brother's son, whether one has to provide சகோதரன் மகன் OR அண்ணன் மகன் OR தம்பி மகன் OR both. The problem is the same in Telugu too.

2. HWN ID 1804. भाई की लड़की /bhai ki ladkhi/ `brother's daughter'

Synsets: भतीजी,भ्रातृजा /BatIjI, BrAtRujA/

This is similar to the one given in 1 above; whether one has to provide சகோதரன் மகள் OR அண்ணன் மகள் OR தம்பி மகள் or both. The problem is the same in Telugu too.

3. HWN ID 683: मामा की लड़की /mAmA kI laDakI/ `uncle's daughter'

Synsets: ममेरी बहन,ममेरी बहिन,मातुलेयी,ममेरी भगिनी

mamerI bahana,mamerI bahina,mAtuleyI,mamerI BaginI

The concept in Telugu can be given as /mAmA kUturu/ But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in case of Telugu, if it is elder daughter then it is /vodhina/ and younger is /maradalu/.

4. HWN ID 2861: मामा का लड़का /mAmA kA laDakA/ `uncle's son'

Synsets: ममेरा भाई,ममियाउत भाई,मातुलेय

mamerA BAI,mamiyAuta BAI,mAtuleya

The concept in Telugu can be given as /mAmA koDukku/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in case of Telugu, if it is elder son then it is /bAva/ and younger is /bAvamaridi/.

5. HWN ID 685: फूफा का लडका /PUPA kA laDakA/ `aunt's son'

Synsets: फुफेरा भाई,फुफेरा भइया,फुफेरा भैया,पितृष्वस्राय

PuPerA BAI,PuPerA BaiyA,PuPerA BaiyA,pitRuShvasrAya

The concept in Telugu can be given as /atta koDukku/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in Telugu, if it is elder son then it is /bAva/ and younger is /bAvamaridi/.

6. HWN ID 686: फूफा की लडकी /PUPA kI laDakI/ `aunt's daughter'

Synsets: फुफेरी बहन,फुफेरी बहिन,फुफेरी भगिनी

PuPerI bahana,PuPerI bahina,PuPerI BaginI

The concept in Telugu can be given as /atta kUturu/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because in Telugu, if it is elder daughter then it is /vodhina/ and younger is /maradalu/.

7. HWN ID 9550: बुआ के पति या पिता के बहनोई /buA ke pati yA pitA ke bahanoI/

`the brother of your father or mother; the husband of your aunt'

Synsets: फूफा /PUPA/

This is an interesting concept in Hindi. In Hindi, the brother of one's father or mother as well as the husband of one's aunt is /PUPA/ but when it comes to Dravidian language they are to be given as different concepts. i.e. `the brother of one's father' as one concept; `the brother of one's mother' as one concept and `the husband of one's aunt' as separate concept. Let us detail these concepts.

a. `brother of one's father' is சித்தப்பா

b. `brother of one's mother' is தாய் மாமன்

c. `the husband of one's aunt' is மாமன்

8. HWN ID 4673: पत्नी का भाई /patnI kA BAI/ `wife's brother'

Synsets: साला, सार, नकलपरवाना, श्वशुर्य

sAIA, sAra, nakalaparavAnA, SvaSurya

The concept in Telugu can be given as /bAriya sOdaruDdu/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in Telugu, if it is elder brother then it is /bAva/ and younger is /bAvamaridi/.

9. HWN ID 6365: पति की बहन /pati kI bahana/ `wife's sister'

Synsets: ननद, ननदी, नंदिनी, नन्दिनी, ननंद, ननन्द, ननदिनी

nanada,nanadI,naMdinI,nandinI,nanaMda,nananda,nanadinI

The concept in Telugu can be given as /barta sOdari/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in Telugu, if it is elder sister then it is /vodhina/ and younger is /maradalu/.

10. HWN ID 7194: वह जो संबंध के विचार से किसी के बहन का पुत्र हो

/vaha jo saMbaMdha ke vicAra se kisI ke bahana kA putra ho/

`... sister's son'

Synsets: भानजा,भांजा,भान्जा,भागिनेय,बहनौता,बहनोत

BAnajA,BAMjA,BAnjA,BAGineya,bahanautA,bahanota

In the above example, the Hindi concept is `in relationship, it is anybody's sister's son'. But if one has to provide an appropriate concept in Telugu, then we have to split the concept into two. That is, if it is one's (male) sister's son, then the synset is /menalludu/ and if it is one's (female) sister's son, then /kodukku/. This is the same in Tamil, viz. மருமகன் and மகள்.

11. HWN ID 7195: संबंध के विचार से किसी के बहन की पुत्री या ननद की पुत्री

saMbaMdha ke vicAra se kisI ke bahana kI putrI yA nanada kI putrI

Synsets: भानजी,भांजी,भाँजी,बहनौती,भागिनेया

BAnajI,BAMjI,BA~MjI,bahanautI,BAGineyA

The Hindi concept is more general, i.e. it does not make a distinction between the speaker, whereas when it comes to Telugu and Tamil, one has to make the distinction. If the speaker is a male, then his sister's daughter is मருமகன், if it is a female, then her sister's daughter is மகள்.

In the above example, only few of the problems faced while constructing in Telugu and Tamil using Hindi concepts are presented. The reason for this is two-fold, viz. first, Hindi concepts are built based on most commonly used words, so one can see how shallow the synsets are and secondly, it is a problem of two different language families in the expansion approach. In Trauttman (1995) words, one needs a hierarchy of constructs, the genetic constructs – Dravidian and Indo-Aryan – and the synthesizing construct of Indian kinship to distinguish the Dravidian data from the non-Dravidian.

Summary

While WordNets are being developed for almost all the major Indian languages, one should give special attention while constructing synsets which is the core of WordNet. As seen from the examples above, it exemplifies that constructing WordNet based on Indo-Aryan languages into Dravidian languages is not a trivial task. It is also clear that kinship relation in Hindi WordNet is shallow and hence one has to take into account different culture specific constructs into consideration. This brings us to some interesting challenges in the construction of WordNets: How to integrate language independent constructs in IndoWordNet; How to handle problems such as male-female distinction in IndoWordNet and vice-versa; elder-younger, distinction which is prominent in Telugu but not in a cognate language such as Tamil? How to represent kinship hierarchy in WordNet? The only solution at this point of time is to build a Domain Ontology within the framework of WordNet Domains.

References

1. Cruse, D.A. 1986. *Lexical Semantics.*, Cambridge: Cambridge University Press
2. Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database.* Cambridge: The MIT press.
3. Forner, Pamela. 2005. *WordNet Domains.* ITC-irst, Povo-Trento, Itala, Document Version 1.0.

4. *Kriyavin taRkaalat tamizh akarati*. 2008. Chennai: CreA.
5. Luisa, B. et.al. 2004. Revising the WordNet Domains Hierarchy: Semantics, Coverage and Balancing. *Coling 04*.
6. Trautmann, Thomas R. 1995. *Dravidian Kinship*. New Delhi: Vistaar Publication.
7. Miller, G. A. 1990. 'WordNet: An Online Lexical Database'. *Special Issue of International Journal of Lexicography*, 3:4.
8. Miller, G.A. 1995. 'WordNet: A Lexical Database for English', *Communications of the ACM*, 38:11, 39-41.
9. Narayan, D., Chakrabarty D., Pandey P. and Bhattacharyya, P. 2002. 'An Experience in Building the Indo WordNet- a WordNet for Hindi', *International Conference on Global WordNet*, Mysore.
10. Nida, E. A. 1975. *Compositional Analysis of Meaning: An Introduction to Semantic Structure*. Mouton: The Hague.
11. Rajendran, S. 2001. *taRkaalat tamizhc coRkaLanjciyam* [Modern Tamil Thesaurus]. Thanjavur: Tamil University Publication.
12. *Tamil Lexicon*. 1982. Madras: University of Madras Publication, Vols.1-6.
13. <http://wordnet.princeton.edu/> (Princeton English WordNet)
14. <http://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php> (Hindi WordNet linked with Indian languages)
15. <http://www.globalwordnet.org> (Global WordNet Association)

கணிணியியலில் தமிழ்ப் பயன்பாடு

இலக்குவனார் திருவள்ளுவன்

7/1, மாவு ஆலை முதல் தெரு

மயிலாப்பூர், சென்னை 600 004

பேசி: 98844 81652 / 044 6499 3317

thiru2050@gmail.com / thiru2050@yahoo.co.in

எத்துறையாயினும் அத்துறையறிவு தாய்மொழியில் வெளிப்படுத்தப் பட்டால்தான் அம்மொழியினருக்கு முழுப் பயன்பாடு கிட்டும்; அத்துறையும் சிறப்பான வளர்ச்சியை எட்டும். அந்த வகையில் கணிணியியலில் முழுமையும் தமிழ்மொழி பயன்படுத்தப்பட்டால்தான் கணிணியியல் முழு வளர்ச்சியடைந்ததாகும்.. இப்பொழுது அந்நிலை இன்மையால், அதனை வலியுறுத்துவதே இக்கட்டுரையின் நோக்கம்.

கட்டுரையாளர்களும் நூலாளர்களும் இதழாளர்களும் தமிழில் கணிணியியலை விளக்குவதில் பெரும் ஆர்வம் காட்டிவருகின்றனர். ஆனால், அவ்வாறு விளக்குவதில் உள்ள ஆர்வம் தமிழைப் பயன்படுத்துவதில் இல்லை. கணிணிக் கலைச்சொற்களாக நல்ல தமிழ்ச் சொற்கள் இருப்பினும் அதைப் பயன்படுத்தாதவர்களும் உளர்; தமிழ்க் கலைச்சொற்கள் இன்மையால் அயல் மொழிச் சொற்களைத் தமிழ் வரிவடிவில் எழுதுவோரும் உளர். எனவே, இந்நிலை மறைந்து இன்னிலை தோன்றப் பின்வரும் செயற்பாடுகளில் ஈடுபட கணிணியியலாளர்கள் முன் வரவேண்டும்.

1. இருக்கின்ற தமிழ்க் கலைச்சொற்களைப் பயன்படுத்தல்.
2. இல்லாதவற்றிற்குப் புதிய கலைச்சொற்களை உருவாக்கல்.
3. நேர் பெயர்ப்புச் சொற்களைத் தவிர்த்தல்
4. ஒலி பெயர்ப்புச் சொற்களை விலக்குதல்
5. நடைமுறையில் பொருந்தாச் சொற்கள் இருப்பின் தக்க கலைச்சொற்களை உருவாக்கல்
6. சொற்சுருக்க எழுத்துகளைத் தமிழில் குறிப்பிடல்.
7. தலைப்பெழுத்துச் சொற்களைத் தமிழில் குறிப்பிடல்.
8. விசைச் சொற்களைத் தமிழில் குறிப்பிடல்.
9. கணிமொழிக் கட்டளைகளைத் தமிழில் அமைத்தல்.
10. கணிப்பொறியின் பகுதிகளைத் தமிழிலேயே குறித்தல்

சுருக்கமாகச் சொல்வதாயின், தமிழை மட்டுமே தெரிந்த ஒருவர் கணிணியியலை நன்கு புரிந்து கொள்ளும் அளவிற்குத் தமிழை மட்டுமே பயன்படுத்திக் கணிணியியலை விளக்கும் காலம் விரைவில் வர வேண்டும்.

தமிழ் ஆர்வலரான கட்டுரையாளர் சிலர், தத்தம் படைப்புகளில் நல்ல தமிழ்ச் சொற்களைக் கையாண்டுவரினும், கணிணித்தமிழ் அறிஞர் சிலர் நல்ல தமிழ்ச்சொற்களைத் தொகுத்து அகராதிகள் வழங்கியிருப்பினும், அவற்றை அறியும் தேடுதல்-வேட்கையின்றியும், அல்லது அறிந்தாலும், அத் தமிழ்ச் சொற்களைப் பயன்படுத்த வேண்டும் என்ற கடப்பாட்டு உணர்வு இல்லாமலும், கணிணித்துறையினர் ஆங்கிலச்சொற்களையே கையாண்டு கணித்தமிழ் வளர்ச்சிக்குத் தடையாக இருக்கின்றனர். எனவே, அறிமுகப் படுத்தப்பட்ட கலைச்சொற்களைப் பயன்படுத்தவேண்டும் என்ற உணர்வு படைப்பாளர்களுக்கு வர வேண்டும்.

கலைச்சொல் பெருக்கத்திற்குத் தடையாக இருப்பது சொல்லைப் புரிந்து கொண்டு படைக்காமல், 'சொல்லுக்குச் சொல்' என்ற நேர்முறையில் ஆக்கப்படும் கலைச்சொற்களும், தமிழ்ச்சொற்களைக் கையாளாமல் ஒலிபெயர்ப்புச் சொற்களாக மூலச்சொற்களைக் கையாளலுமாகும். இவற்றை உணர்ந்து, புத்தம்புதுக் கலைச் சொற்களை நாளும் உருவாக்கவும், உருவாக்கப்பட்ட கலைச் சொற்களைப் பயன்படுத்தவும் நாம் முன்வர வேண்டும். கலைச்சொற்கள் சுருங்கியனவாகவும், அவற்றின் அடிப்படையில் மேலும் புதிய கலைச்சொற்களை ஆக்க வாயிலாகவும் அமைய வேண்டும்.

அறிவியல் துறைகளைப் புரியவைப்பதற்கும் அறிந்துகொள்வதற்கும் கையாளப்படும் கலைச்சொற்கள் தன்-விளக்கமாயும் எளிமையாயும் அமைய வேண்டும். அவ்வாறு இல்லாச் சூழலில், தவறாகப் புரிந்து கொள்ளவோ, விளங்காமல் குழப்பம் அடையவோ வாய்ப்புகள் ஏற்படுகின்றன. எனவே, விரைந்து வளரும் கணினியியலில் துறைவளர்ச்சிக்கேற்ற கலைச்சொல் பெருக்கமும் அமைய வேண்டும். கலைச்சொற்கள் பெருகுவதற்கான தடைகளை நீக்க,

1. ஒவ்வொருவர் ஒவ்வொரு வகையாகக் கையாளுதல்.
2. சில நேரங்களில் ஒருவரே வெவ்வேறு வகையாகக் கையாளுதல்.
3. நடைமுறைக்கு நல்லசொற்கள் வந்துவிட்டபின்னும் கொச்சையாகக் கையாளுதல்.
4. சுருங்கிய கலைச்சொல்லாக இல்லாமல், விளக்கச் சொற்றொடராகக் கையாளுதல்
5. பொருள்விளக்கமான கலைச்சொல்லைக் கையாளாமல், நேருக்குநேர் மொழி பெயர்த்துக் கையாளுதல்.
6. தவறான சொல்லாக்கத்தைக் கையாளுதல்
7. சொல்லும் அதன் பயன்பாட்டுக் காலத்திற்கு ஏற்பப் பொருள்மாற்றம் அடைகிறது. எனவே, இதுதான் இச்சொல்லுக்குப் பொருள் என்னும் பிடிவாதம் இன்றிச் சூழலுக்கேற்ற பொருள் விளக்கத்தைக் கையாளாமையாகியவற்றை அறவே நீக்குதல் வேண்டும்.

ஆதலின், நடைமுறையில் இல்லாத கலைச்சொற்களுக்குத் தேடுதல் வேட்கையுடன் புதிய கலைச்சொற்களை உருவாக்க வேண்டும். கலைச்சொல் புனையும் ஈடுபாடும் ஆர்வமும் இல்லாதவர்கள் கலைச்சொல் வல்லுநர்கள் மூலம் புதிய கலைச்சொற்களைப் படைக்கத் தூண்டதலாய் இருத்தல் வேண்டும். அறிமுகமாகியுள்ள கலைச்சொற்கள் உரிய பொருள் தராதனவாகவும் தொடர்போன்றும் அமைந்து இருப்பின், அவற்றிற்கும் உரிய பொருத்தமான சுருக்கமான கலைச்சொற்களை உருவாக்கிப் பயன்படுத்த வேண்டும். ஒவ்வொருவர் ஒவ்வொரு வகையான கலைச்சொற்களைக் கையாளுதலும், ஒருவரே வெவ்வேறுிடத்தில் வெவ்வேறு வகையான சொற்களைக் கையாளுதலும் படிப்பவர்களிடையே குழப்பத்தை ஏற்படுத்தி எதிர் விளைவுகளை உருவாக்கும். எனவே, நிலைத்துவிட்ட நல்ல சொற்களை மாற்றும் முயற்சியைக் கைவிட வேண்டும். அதே நேரம் நடைமுறையிலுள்ள சொல்லைவிடப் பொருத்தமான கலைச்சொல் அறிமுகப்படுத்தப்பட்டால் பிடிவாதத்துடன் முந்தைய சொல்லையே கையாளாமல் புதிய கலைச் சொற்களைக் கையாளும் மனப்பக்குவமும் வேண்டும். ஓரிடத்தில் பொருத்தமாக உள்ள கலைச்சொல் வேறுிடத்தில் உரிய பொருளைத் தராமல் பொருந்தாமல் நிற்கும். எனவே, சொல் இடத்திற்கேற்ற பொருளைப் பெறும் என்பதை உணர்ந்து சூழலுக்கேற்ற கலைச்சொல்லையே பயன்படுத்த வேண்டும். தேவையான இடங்களில் அடைப்பிற்குள் ஆங்கிலச்சொல்லையோ நடைமுறையில் உள்ள சொல்லையோ குறிப்பிடத் தயங்கக் கூடாது.

மூலச் சொற்களையும் தலைப்பெழுத்துச் சொற்களையும் சுருக்க அமைப்புச் சொற்களையும் ஆங்கிலத்திலேயே குறிப்பிட்டால் தவறல்ல என்னும் மனப் போக்கு பெரும்பாலாரிடம் உள்ளது. இதுவும் தவறான நிலைப்பாடாகும். இவையும் தமிழில் இருக்கும்பொழுது கணினியறிவியல் மேலும் எளிமையாகத்

திகழும். ஐ.நா. என்பது போன்ற தமிழ்ச் சுருக்கக் குறியீடுகள் பொருளை விளங்க வைக்க உதவுவதை எடுத்துக்காட்டாகக் கூறலாம். தமிழில் இருந்தால் புரியாது என்று சொல்வதெல்லாம் மேலோட்டச் சிந்தனையே!

கணினியியலில் ஆங்கில ஒலிபெயர்ப்பிலேயே கலைச் சொற்களும் தலைப்பெழுத்துச் சொற்களும் எண்ணிலடங்கா அளவு கையாளப்பட்டுத் தமிழ் மொழி சிதைந்து வருவதைப் பலரும் உணரவில்லை. 'மணிப்பிரவாளம்' என்ற பெயரில் மொழிக்கொலை புரிந்து பாழ்பட்ட நிலையிலிருந்து அண்மைக் காலத்தில் மீண்டுவரும் வேளையில் ஆங்கிலக்கலப்பு விளைவிக்கும் தீங்கைப் பெரும்பான்மையர் புரிந்து கொள்ளவில்லை. பிற அறிவியல் துறைகளில் நிகழும் சொல்லாக்கத் தவறுகள்தாம் கணினியியலிலும் நடைபெறுகின்றன. ஆனால், பிற துறைகளுடன் ஒப்பிட முடியாத அளவு கணினியியலில்தான் ஆங்கில ஒலிபெயர்ப்புச்சொற்கள் மிகுதியாகக் கையாளப்படுகின்றன. இவை முற்றிலும் உடனடியாகக் களையப்பட வேண்டும். சுருக்கக் குறியீடுகள், தலைப்பெழுத்துகள் என எந்த வடிவிலும் ஆங்கிலத்தைப் பயன்படுத்தாமல் சீன மொழியிலேயே குறிக்க வேண்டும் எனச் சீன அரசு ஆணை பிறப்பித்து நடைமுறைப்படுத்தி வருகிறது. இது போல் தமிழ்நாட்டரசும் ஆணை பிறப்பித்து நடைமுறைப்படுத்த வேண்டும்.

சொல்லில் உயர்வு தமிழ்ச் சொல்லே என்னும் பாரதியாரின் பொன்மொழியை உணர்ந்து, தமிழில் எண்ணித் தமிழிலேயே எழுதத் தொடங்கினால் அரிய கலைச் சொற்களைக்கூட அழகுதமிழில் அருமையாகக் கூற இயலும். தமிழ் எழுத்துகளில் அமைந்தன மட்டுமே தமிழ் என்பது நம் முன்னோர் கூற்று. ஆகவே, தமிழ்ப்படைப்புகளில் அயற்சொற்களும் கிரந்த எழுத்து முதலான அயல்எழுத்துகளும் பயன்படுத்தக்கூடா. இவற்றை ஊக்கப்படுத்துவதற்காக அரசு, தமிழ்க்கலைச்சொற்களைப் பயன்படுத்தும் நூல்களை மட்டுமே பாட நூல்களாக வைக்கவேண்டும்; கலப்பு நடையைக் கைவிட்டு நல்ல தமிழில் எழுதப்படும் நூல்களுக்கு மட்டுமே பரிசுகள் வழங்க வேண்டும். தமிழ்ப்பகைவர்களுக்குப் பட்டங்களும் விருதுகளும் பொற்கிழிகளும் வழங்கி மொழி இனஅழிப்பிற்குத் துணை போகாமல் தமிழ் அன்பர்களை மதித்துப் போற்ற வேண்டும்.

கலைச்சொற்களை மட்டும் தமிழில் வழங்கினால் போதுமா? கணிக்கட்டளைகளையும் தமிழிலேயே அமைத்தல் வேண்டும். அதற்கு முதற்கட்டமாகக் கணினிச் செயற்பாட்டுக் கட்டளைகளைக் குறிப்பிடும் விசைகளின் பெயர்கள் பின்வருவன போல் தமிழில் இருக்க வேண்டும்.

- | | |
|---------------------------------|--------------------------------------|
| · Enter Key - புகுவி விசை | · Page up Key - ஏற்றி விசை |
| · Control Key - யாப்பு விசை | · Page down Key - இறக்கி விசை |
| · Alternate Key - வினை விசை | · Pause Key - நிறுத்தி விசை |
| · Delete Key - நீக்கி விசை | · Print Screen Key - பதிப்பி விசை |
| · Escape Key - விலக்கி விசை | · Up Arrow Key - மேலம்பு விசை |
| · Home Key - ஆதி விசை | · Down Arrow Key - கீழம்பு விசை |
| · End Key - அற்றவிசை | · Left Arrow Key - இட அம்பு விசை |
| · Shift Key - முறைமை விசை | · Right Arrow Key - வல அம்பு விசை |
| · Tab Key - பெயர்த்தி விசை | · Back Space Key - முன்னிட விசை |
| · Number Lock key- எண்தாழ் விசை | · Functional Keys - செயல் விசைகள் |
| · Scroll Lock Key - சுருணை விசை | · User Keys - பயனர் விசை |
| · Insert Key - செருகி விசை | · Caps.lock key - முறைமைத் தாழ் விசை |

இவை போன்று கட்டளைச் சொற்களையும் தமிழில் அமைத்து இம்முயற்சியை விரைவுபடுத்த வேண்டும் கணிப்பொறியின் பகுதிகளைத் தமிழிலேயே குறித்தல் வேண்டும் அப்பொழுதுதான் கணியியல் குறித்த முழுமையான தமிழ்நூல்களைப் படைக்க இயலும்.

இவையனைத்தையும், தமிழில் அமைக்கக் கணினியியலாளர்கள் முன்வரின் கணினியியலில் தமிழ் தலைமையுற்றுத் திகழும். தமிழ்வழியாகக் கல்வி அமையாமையாலேயே நம் நாட்டில் புதிய புனையும் அறிஞர்களும் கண்டுபிடிப்பாளர்களும் உருவாகவில்லை என்பார் செந்தமிழ்ச் செம்மல் பேராசிரியர் சி.இலக்குவனார். கணினி உலகில் நாளும் அறிஞர்கள் பெருக வாழும் மொழியாம் தமிழில் முழுமையாய் கணினியறிவியல் அமைய வேண்டும்.

செயல் செய்வாய் தமிழுக்குத்

துறைதோறும் துறைதோறும் சீறிவந்தே

என்னும் பாவேந்தர் பாரதிதாசன் கட்டளைக்கிணங்க நாம் கணினியறிவியலிலும் தமிழ்ப்பயன்பாட்டை முழுமையாகக் கொண்டு வர வேண்டும். அதுவே நாம் செய்யும் எப்பணிக்கும் முதற்பணியாய் அமைதல் வேண்டும்.

அனைத்திலும் தமிழ்!

கணியறிவியலிலும் தமிழ்!

பார்வைக்குரியன

கட்டுரையாளரின் படைப்புகள்

1. ஒரு சொல் - பல பொருள் : கலைச்சொல்லாக்க வளர்ச்சியின் முட்டுக்கட்டை (உலகத் தமிழ் மாநாட்டுக் கட்டுரை, மலேசியா)
2. இதழியல் சொல்லாக்கம் - திறனாய்வும் நெறிமுறையும் (உலகத் தமிழ் மாநாட்டுக் கட்டுரை, மலேசியா)
3. கணினிக் கலைச்சொற்கள் (மதுரை காமராசர் பல்கலைக் கழக வியாழன் வட்டக் கட்டுரை)
4. இன்றைய தேவை குறுஞ்சொற்களே (உலகத் தமிழ் மாநாட்டுக் கட்டுரை, தஞ்சாவூர்)
5. அன்றாட நடைமுறையில் சொல்லாக்கம்
6. கணினியியலில் நேர்பெயர்ப்புச் சொற்களும் ஒலிபெயர்ப்புச் சொற்களும் (ஐந்தாவது இணையத்தமிழ் மாநாட்டுக் கட்டுரை, அக்டோபர் 2009, செருமனி)
7. கணிவிசைப் பெயர்கள் (சென்னைப் பல்கலைக்கழகத் தமிழ்த்துறைக் கருத்தரங்கக் கட்டுரை)

பிற

8. Computer Dictionary (English - Tamil) - இராம்குமார்; தெ.சை.சி.நூற்பதிப்புக்கழகம்
9. The Illustrated Computer Dictionary (Third Edition): Donald D.Spencer ; Universal Book Stall
10. கணிப்பொறிக் கலைச் சொல் அகராதி : வளர்தமிழ் மன்றம், அண்ணா

பல்கலைக்கழகம்

11. அறிவியல் அகராதி : பேராசிரியர் அ.கி.மூர்த்தி : மணிவாசகர் பதிப்பகம்
12. 'தமிழ் கம்ப்யூட்டர்' முதலான இதழ்கள்

Role of Regular expression (RE) in Morphological Analysis

R.Shanmugam
Madras University
shanfaace@gmail.com

Abstract

The aim of this article is to analyze the role of Regular Expression in Tamil Morphological Analysis. Morphological Analysis is essential for Natural Language Processing (NLP) and Machine Translation (MT). In morphological analysis, what we do is to parse the inflected word into root and affixes and then to tag them for grammatical categories. To build a Syntactic Parser, we need a Morphological Parser with POS (Parts-of-speech) tagger. The input for the Syntactic Parser is the output of the Morphological Parser. There are many formalisms and tools are used in the field of Morphological Analysis. Regular expression is one among the best tools for Morphological analysis.

Introduction

Regular expression is the standard notation for characterizing strings (combination of characters). It is a formula in a special language for specifying simple classes of strings. Formally it is an algebraic notation for characterizing strings. Regular expression was introduced by Kleene (1956). A string is any sequence of characters like letters, numbers, spaces, tabs, punctuation. Space which is also a character because it has encoding value. Regular expression needs a pattern (search type) to search strings. The following table shows how the words are matched by regular expression.

Regular Expression	Example pattern matched
/puththakam/	avaN <i>puththakam</i> patiththaaN
/kalvi/	<i>kalvi</i> aNaivarukkum avaciyam
/niir/	kutikka <i>niir</i> veeNntum

So the regular expression /puththakam/ matches any string containing the substring *puththakam* like the above example. Here the slashes (around the pattern) are used to clarify what is regular expression and what is pattern. (This notation used in perl). The upcoming table shows some of the important Regular Expression symbols which are used in Tamil morphological parsing.

Patterns in Regular expression: (Based on Perl language)

RE symbol, Name	Example pattern	Match
\~, tint	marankkaL =~ /kaL/	Pattern matching symbol finds if 'kaL' is the substring of marankkaL
\(), Open and	/marank(kaL ai)/	Whether marankkaL or marankkaLai

close parenthesis		
\ , pipe line	\$a =~/maram maNithaN/	Whether \$a is maram or maNithaN
\?, Question mark	karuththuk?kaL	karuththkaL or karuththukkaL
\\$, scalar	avarkaL =~/kaL\$/	Whether 'avarkaL' ended with kaL
+,	ceythaaN =~/th+/ patiththaaN =~/th+/	One or more 'th' in ceythaaN One or more 'th' in patiththaaN

Levels in Morphological Analysis

There are three levels involved in Morphological Analysis such as 1. Root checking, 2. Affix Stripping, 3. Morphophonemic changes. We can use Regular Expression for the above three levels. It has much influence in Morphological Analysis. It is a very handy tool to make this kind of Morphological analysis. It can be implemented easily with Perl language.

The Role of Regular expression in Root checking

This is a basic task in Morphological Analysis. Under this task, the root word existence in the database would be found. The following code may useful to know the importance of RE in this task.

```

open(filehandle,"C:\\DataBase\\Noun.txt") || die;
while ($line=<filehandle>)
{
if($line =~ m/\b$input\b/)
{
$root word=$&;
last;
}
}

```

Here the line in bold is important. The character 'm' is stands for matching i.e., to verify whether the database word is same with the end user input. '\b' is used to denote the word boundary. '\$&' is called register in RE it is used to store the matched word into the variable \$root word.

The Role of Regular expression in Affix Stripping

Affix stripping is used to strip the affixes from the given input. The following sample coding explains the role of RE in Affix Stripping section.

```

Stripping tense suffix from the input stem "patithth"
if($input =~ /([v | N | t | R])$ | (pp?)$ | (iN)$ | ((n)?(th)+)$ | (kk?iN?R)$/)
{
$tense=$&;
$root=substr($input,0,-length($tense));
}

```

The above code is having 14 tense forms for stripping purpose. It is a great strength of RE. We can implement this in one line and it gives the following output.

```
$input (Input variable) = patithth  
$tense = thth  
$root = pati
```

The Role of Regular expression in Morphophonemic Section

This section is playing the key role for this programming. This section includes three sub-functions such as addition, deletion, substitution. The three functions would be used to make the remaining stem into root. If we want to change the stem 'marathth' into maram the following code is needed.

```
$input = marathth  
if(($input=~m/thth$/) and ($case ))  
{  
$input=~s/thth/m/;  
}
```

Output: \$input = maram

Here the bold lines are showing the importance of RE. 'if' condition is used to identify the occurrence of 'thth' in input and substitution is used to change the stem 'marathth' into root 'maram'.

Conclusion

Regular expression is a handy formalism to make Morphological Parser and it is common for languages we can adopt these ideas for other languages too, so that if we have a clear view on this we will make a good Morphological Analyzer for our language.

கணினிக் கலைச்சொல்லாக்கத்தில் புதுச்சொல் உருவாக்கம்

லா. சார்லஸ், எம்.ஏ., எம்.பில்.,

தமிழ் விரிவுரையாளர்,

தூய வளனார் தன்னாட்சிக் கல்லூரி, திருச்சிராப்பள்ளி-02.

அலைபேசி:9842599311

charlesjc85@gmail.com

மொழி என்பது அனைத்து நாடுகளிலும் உயிர் நாடியாகக் கருதப்பட்டு வருகின்றது. ஒரு மொழிக்கு உயர்வு கற்பிக்க வேண்டும் என்று கருதினால் அக்காலத்தில் அந்தமொழியைத் தேவமொழி என்று கூறினர். இன்று ஒரு மொழியை உயர்வுபடுத்த வேண்டும் என்றால் அம்மொழியை அறிவியல் தொழில் நுட்பத்திற்கு ஏற்ற மொழியாக ஆக்குவதே அந்த மொழிக்குச் செய்யும் தொண்டு ஆகும். அறிவியல் தொழில் நுட்பத்திற்கு ஏற்ற மொழியாக ஒரு மொழி விளங்க வேண்டும் என்றால் அம்மொழியில் கலைச் சொற்கள் உருவாக்கப்பட வேண்டும்.

மொழி முதன்மை பெறாத இடத்தில் சிந்தனை செழித்து வளர்வது இல்லை. ஜப்பான் போன்ற வளர்ந்த நாடுகளில், தாய்மொழி வழிக் கல்வியினையும், அதனால் அவர்கள் பெற்று வருகின்ற உயர்வினையும் உலகமே இன்று கண்டு வியக்கின்றது. தாய்மொழியால் இயலாததை எம்மொழிகளாலும் சாதிக்க முடியாது என்பதைப் பாரதி முதல், பல்வேறு அறிஞர்களும் கூறி வருகின்றனர். அதன் பிரதிபலிப்பே கலைச்சொற்கள் இன்று பல துறைகளில் செழிக்கக் காரணமாக அமைகின்றது. அவற்றில் ஒன்று தான் தற்காலத்தில் நாள்தோறும் வளர்ந்து வரும் கணினித் துறையாகும். எங்கும் கணினி எதிலும் கணினி எனப் பல்வேறு பரிமாணங்களில் கணினியின் பயன்பாடு உணரப்பட்டு வருகிறது. அந்த வகையில் கணினிக் கலைச் சொற்களைத் தமிழிலே ஆக்கம் செய்வது காலத்தின் கட்டாயம் ஆகும். தமிழ்மொழியில் இத்தகைய கணினிக் கலைச் சொற்களின் உருவாக்கம் குறித்தும், அதன் வாயிலாகத் தமிழ்மொழி பெறும் வளர்ச்சி குறித்தும் ஆராய்வது இவ்வூய்வின் நோக்கங்களாக அமைகின்றன.

கலைச்சொல்லாக்கம் - வரையறை

கலைச் சொல்லியலின் தந்தை பேராசிரியர் யூஜின் யூஸ்டர் கலைச் சொல் பற்றிக் கூறும்போது, “ஒரு துறையின் கலைச் சொல்லானது, அத்துறையின் பயன்பாட்டுக்கு மட்டுமின்றி, சமூக வாழ்வியலில் உள்ள பல தரப்பட்ட மக்களுக்கும் புரிந்துணர்வனவாக அமைதல் வேண்டும்” (நா.ஜானகிராமன் (2003):15) என்று குறிப்பிடுகின்றார். அதனடிப்படையில் கலைச் சொற்களின் முழுமுதல் காரணமே அவரவர்கள் தங்கள் தாய்மொழி வாயிலாகக் கல்வி பயில வேண்டும் என்பதே ஆகும்.

“ஒரு மொழிக் கூறிலுள்ள மையத்தினை மற்ற மொழிக்கு மாற்றுவதும், அம்மொழிக் கூறின் மையக் கருத்திற்கு ஏற்றதொரு சொல்லைப் பிறிதொரு மொழியில் ஆக்கிக் கொள்வதும் கலைச் சொல்லாகும்” (இராதாசெல்லப்பன் (2006):126) என்கிறார். அத்தகைய கலைச் சொல்லாக்கத்தை படைப்பாக்கம் என்பர்.

சொல்லாக்கம், மொழியாக்கம், படைப்புச் சொற்கள், கலைச் சொற்கள் இவை அனைத்தும் புதுச் சொல் படைத்தலின் மூலங்கள் ஆகும்;. இவையே தமிழ் மொழியின் வாயிலாகக் கணினித் துறைக்குத் தேவையாய் உள்ளன. எளிமையாக்கத்திற்கும், நீண்ட கருத்துரையாடல்களை சுருக்குவதற்கும், கருத்துப்புலப்பாட்டிற்கும், சொல்லாக்கம் தேவையாய் உள்ளது. எனவே கணினித் துறையும்,

தமிழ்த்துறையும் இணைந்து வளரும் போது, முறையான கலைச்சொல் பணியும் கருத்துத் தொடர்புகளைத் தெளிவாக்கும் பணியும் சிறப்படையும்.

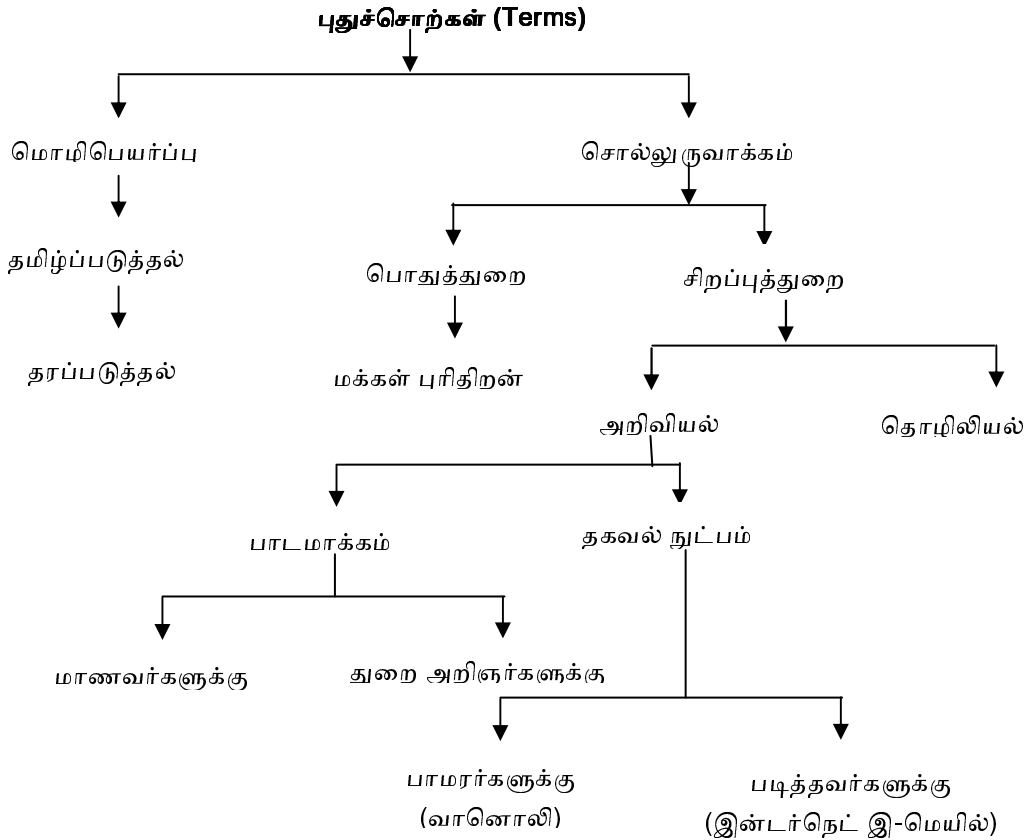
கலைச்சொல்லாக்க வகைகள்

கணினிக் கலைச் சொல்லாக்கம் என்பது மூன்று வகைகளில் அமைகின்றது. அவை 1) ஒலிபெயர்ப்பு 2) மொழி பெயர்ப்பு 3) புதுச்சொல் உருவாக்கம் என்பவை ஆகும்.

சிதம்பரநாதச்செட்டியார்

- வேற்றுமொழிச் சொற்களைத் தமிழ்ப்படுத்தி ஏற்றுக்கொள்வதை ஒலிபெயர்த்தல் என்றும்,
- வேற்றுமொழிச் சொல்லுக்குச் சரியான பொருளைக் கண்டு அதற்கு ஈடான ஒரு சொல்லைத் தமிழில் காணல் அல்லது ஆக்கிக் கொள்ளலை மொழிபெயர்த்தல் மற்றும் புதுச்சொற்கள் படைத்தல் என்றும்,
- புதிதாக வேண்டியிருக்கும் சொல்லுக்கு நெருங்கிய பொருளுடைய ஒரு வழக்கிழந்த சொல்லை எடுத்து அதற்குப் புதிய பொருளைக் கொடுத்தல் புதுப்பொருள் கொடுத்தல் (சிதம்பரநாதச்செட்டியார் (1957):28) என்றும் விளக்கம் அளிக்கின்றார்.

உலகின் பல நாடுகளிலும் தத்தம் மொழியில் கலைச் சொற்கள் ஆக்குபவர் இந்த மூன்று முறைகளையே பின்பற்றுகின்றனர். ஒவ்வொரு முறையிலும் சில நன்மைகளும், சில தீமைகளும் இருப்பினும் இந்த மூன்று முறைகளையும் பின்பற்ற வேண்டியுள்ளது.



புதுச்சொல் உருவாக்கம்

அறிவியல் வளர்ச்சியின் வேகத்திற்கு ஈடுகொடுக்க மொழிகள் போராடிக் கொண்டிருக்கும் காலம் இது. எல்லா மொழிகளும் காலத்திற்குக் காலம் மாறி வந்துள்ளன என்பது மறுக்க முடியாத ஓர் உண்மையாகும். தமிழ் மொழியின் அமைப்பு பண்டைக்காலத்தில் இருந்தது போல் தற்காலத்தில் இல்லை. காலந்தோறும் மாறும் மொழிப் பண்பிற்கேற்பத் தமிழிலும் ஒலிவடிவம், வரிவடிவம், ஒலியமைப்பு, இலக்கணம், சொற்றொடர், சொல், சொற்பொருள் ஆகியன பல நிலைகளில் பலவித மாறுதல்களுக்கு உட்பட்டுள்ளன.

இவற்றில் புதுச்சொல் படைப்புமுறையானது, என்ற நிலைகளில் அமைவதாக (நா. ஜானகிராமன் (2003):13) குறிப்பிடுகின்றார். புதுச்சொற்களானது மொழிபெயர்ப்பு, சொல்லுருவாக்கம் என்ற இரண்டு நிலைகளில் அமைகின்றன. சொல்லுருவாக்கம் என்பதை ஆய்வாளர் புதுச்சொல் உருவாக்கம் எனக் கையாள்கிறார்.

புதுச்சொல்லுருவாக்கம் என்பது பொதுத்துறை, சிறப்புத்துறை என்ற இரண்டு நிலைகளில் அமைகிறது. பொதுத்துறையானது மக்களுக்குப் புரியும் திறன் அடிப்படையிலும், சிறப்புத்துறையானது அறிவியல், தொழிலியல் என்ற இரண்டு நிலைகளிலும் அமைகின்றன. அறிவியல் நிலையிலிருந்து புதுச்சொற்களானது மாணவர்களுக்கும், துறை அறிஞர்களுக்கும் பாடமாக்கம் செய்வதாக அமைகிறது. தகவல் தொழில்நுட்ப நிலையிலிருந்து புதுச்சொல் உருவாக்கமானது பாமரர்களுக்கும், படித்தவர்களுக்கும் உரிய பயன்பாட்டின் அடிப்படையில் அமைகிறது.

புதுச்சொல்லாக்க உத்திகள்

புதுச்சொல் படைப்பு என்ற சொல் நான்கு வகையான சொல்லாக்க உத்திகளைக் குறிக்கிறது.

- 1) தமிழ்ச் சொல்லாக்க விசுவகளை, ஏற்கனவே வழக்கில் இருக்கும் அடிச்சொற்களுடன் இணைத்தும், தொகைச் சொற்களை உருவாக்கியும் புதிய சொற்களை ஆக்கிக் கொள்ளுதல்.
- 2) ஏற்கனவே வழக்கிலிருக்கும் சாதாரணச் சொல்லைக் கலைச் சொல்லாகப் பொருத்திக் கொள்ளுதல்.
- 3) வழக்கிழந்து போன சொற்களுக்குப் புதுப்பொருள் கொடுத்துக் கலைச்சொற்களை ஆக்கிக் கொள்ளுதல்.
- 4) புதிய சொல்லாக்க முறைகளைக் கையாண்டு புதியதாகவே சொற்களைப் படைத்துக் கொள்ளுதல்.

என (இராதாசெல்லப்பன் (2006):130) அமைகின்றன.

(1). சொல்லாக்க விசுவகளை அடிச்சொற்களுடன் இணைத்துப் புதுச்சொல் படைத்தல் உத்தி முறையாக உள்ளது. அம், இ, கை, மை, பு, சி, அல், முதலிய பல விசுவிகள் சொற்களை ஆக்கிக் கொள்ள உதவுகின்றன.

சான்றாக,

அம்	-	Internet	-	இணையம் (இணை+அம்)
இ	-	Mouse	-	சுட்டி (சுட்டு+இ)
மம்	-	Digital	-	எண்மம் (எண்+மம்)

அவ்வாறே மொழியியல்படி use என்னும் அடிச்சொல் er என்னும் 'செயலைச் செய்பவர்' என்னும் பொருள் அமைந்த விசுவியோடு சேர்ந்து use + er = user என அமைகிறது. இதனையொட்டி user என்பது பயன் + ஆளர் = பயனாளர் என்று கலைச்சொல்லாக உருப்பெறுகிறது.

(2). Disk என்பதற்கு இணையாக 'வட்டு' என்னும் சொல் பொது வழக்கில் கையாளப்பட்டு வருகிறது. (சான்று : Disk Throwing - வட்டு எறிதல்) Floppy Disk என்பதைக் குறிக்க நெகிழ்வட்டு என்னும்

சொல் பயன்படுகிறது. ஆனால் Disk என்னும் அதே சொல் Compact Disk (CD) என்பதைக் குறிக்கையில் 'குறுந்தகடு' என்றும் Hard Disk என்பதைக் குறிக்கும்போது 'வந்தகடு' என்றும் குறிக்கப்படுகிறது. எனவே Disk என்பதற்கு இணையாக 'தகடு' என்னும் வேறொரு வழக்குச் சொல்லும் கையாளப்படுகிறது.

(3). Control unit - கட்டுப்பாட்டு அலகு எனச் சொல்லுக்குச் சொல் மொழி பெயர்க்கலாம். இதனைக் கட்டுப்பாட்டுப் பகுதி என்று மொழிபெயர்ப்பதைக் காட்டிலும் கட்டுப்பாடு + அகம் = கட்டுப்பாட்டகம் எனக் கையாளவது சிறப்பாக அமையும். இதனை அடியொற்றியே CPU - மையச் செயலகம், Memory Unit - நினைவகம் ஆகிய சொல் வழக்குகளும் உருவாக்கப்படுகின்றன. சங்க இலக்கியங்களில் கையாளக்கூடிய அகம் என்னும் சொல் உள்ளத்தையும், புறம் என்னும் சொல் வெளிப்புறத்தையும் குறிப்பதாக அமைகின்றன. இங்கு அகம் என்னும் சங்க கால வழக்கு இடத்தைச் சுட்டும் பொருட்டுக் கையாளப்படுகிறது.

(4). LAN - Local Area Network என்பதை உள்வட்டார இணையம் என்று குறிப்பிடுவதைக் காட்டிலும் 'குறும்பரப்பு வலையமைப்பு' என்று குறிப்பிடுவது புதுப்படைப்பாக்கமாகவே அமைகிறது. இதனை அடியொற்றி WAN - நெடும்பரப்பு வலையமைப்பு எனக் குறிப்பிடுதல் சிறப்பான முறையாக அமைகிறது.

மொழியின் வளர்ச்சிக்கு இன்றியமையாமை

தமிழில் கலைச் சொற்கள் தோன்றிய ஆரம்ப நிலை மிகவும் சிறப்பிற்குரியது. பிறமொழி நூலிலும், வழக்கிலும் உள்ளதைத் தமிழின் மொழிக் கூறுகளுக்கு ஏற்ற வண்ணம் மொழியாக்கம் செய்து புதிய சொல் ஒன்றினை ஏற்படுத்திக் கொள்வதுதான் இன்றைய பெரும்பான்மை நிலை ஆகும். தாய்மொழி வழிப்பாடங்கள் தமிழில் இடம்பெற கலைச் சொற்கள் அவசியத்தை ஏற்படுத்தின.

கணினிக் கலைச் சொற்கள் ஒவ்வொன்றும் பயன்பாட்டிற்கு ஏற்ப மேலும் செழுமை பெறுகின்றன. Chat என்பது சொல்லாடல், விவாதம், அரட்டை என்னும் மூன்று சொற்களாய்ப் பயன்படுத்தப்படுகின்றன. ஆனால் இதற்கு இணையான தமிழ்ச்சொல் 'சொல்லாடல்' என்கிறார் (பெ.மாதையன்(2009):56).

Laptop Computer என்பதை மடிக்கணினி என்று கூறுகின்றோம். இது பயன்பாட்டு அடிப்படையில் அமைந்த எளிய சொல் வழக்கு ஆகும். Software என்பதை மென்பொருள் என்றும், Hardware என்பதை வன்பொருள் என்றும் கூறுகின்றோம். சிலர் இதனை முறையே மென்மம், வன்மம் எனக் குறிப்பிடுகின்றனர். ஆனால் 'வன்மம்' என்னும் சொல்லுக்குத் தமிழில் 'நெடும்பகை பாராட்டல்' என்னும் பொருளே வழங்கி வருவதால் அத்தகைய சொல்லை இங்குக் கையாளவதைத் தவிர்த்து ஏற்கனவே வழங்கி வந்த 'வன்பொருள்' என்பதையும், அதற்கு ஒத்த விதத்தில் 'Software'க்கு மென்பொருள் என்பதையும் வழங்குவது பொருள் புரிவதற்கு எளிமையாய் அமையும். இவ்வாறு ஒரு கணினிச் சொல்லுக்கு பல்வேறு வகையான சொற்கள் கலைச் சொற்களாக அமைவதைக் காணும்போது தமிழ் மொழியின் வளர்ச்சி நிலை மேலோங்கி இருப்பதை அறிகின்றோம்.

பயன்பாட்டின் மையம்

கணினிக் கலைச் சொல்லாக்கத்தில் சொல் படைப்புகள் ஒவ்வொன்றும் பண்டைக்கால பயன்பாட்டில் இருந்து வழக்கு இழந்தவையாக இருந்தாலும் கால வளர்ச்சிக்கு ஏற்ற புதுச் சொற்களாக இருந்தாலும் எந்த அளவிற்கு மக்களின் வாழ்க்கைப் பயன்பாட்டில் இடம்பெறுகின்றதோ அப்போதே அச்சொற்கள் அனைத்தும் வெற்றி பெற்றவையாகக் கருதப்படுகின்றன.

சொல் உருவாக்கமானது பொது மக்களுக்கான பயன்பாட்டுக்கு எளிமை, தெளிவு, தனிமனித நோக்கு, பொதுச் சமுதாயம் என்ற நிலைகளில் அமைதல் வேண்டும். அதே போன்று துறை வல்லுனர்களுக்கான பயன்பாட்டுக்குப் பொதுச்சொல், வட்டாரத்தை ஒலித்தல், மொழி மாற்றம், எழுத்து மாற்றம் என்ற

நிலைகளில் அமைதல் வேண்டும். இவை இரண்டும் சேர்ந்து உலகத் தமிழ் மக்களுக்கான பயன்பாடு என்னும் இலக்கைக் கொண்டு அமைய வேண்டும்.

சான்றாக, Mouse என்ற சொல் சுட்டெலி, சொடுக்கி, சுட்டி என்று பல்வேறு நிலைகளில் சொல் மாற்றம் பெறுகின்றது. 'சுட்டெலி' என்பது அறிமுக நிலையில் அதன் வடிவத்தைப் பொறுத்து வழங்கி வந்த சொல். 'சொடுக்கி' என்பது அந்தச் சாதனத்தின் மீது நிகழ்த்தப்படுகிற 'சொடுக்கு' என்னும் வினையின் அடிப்படையில் உருவான சொல். 'சுட்டி' என்பது அக்கருவியின் பயன் அடிப்படையில் உருவான சொல் ஆகும். மேற்கண்ட மூன்று நிலைகளில் எளிமையாகவும் பொருள் பொருத்தமுடையதாகவும் அமைகின்ற சொல்லாக 'சுட்டி' என்பதையே கருதமுடிகிறது. மேலும் Website என்ற சொல்லிற்கு தகவலகம், மின்னலிடம், வலைக்களம் எனப் பல சொற்கள் வழங்கி வந்தாலும் 'வலையகம்' என்பதே எளிமையாகவும், தெளிவாகவும் விளக்கும் வகையில் பயன்பாட்டின் மையமாய் அமைகின்றது.

சொல்லுருவாக்கத்தில் சிக்கனமுறை

எதிர்காலவியல் நோக்கில் ஒரு பொருளாக இருப்பினும், பொருளுக்கான பெயராக இருப்பினும் எதுவாயினும் சுருக்கி அமைதலே இன்று சிறப்பிற்குரியதாய் உள்ளது. அதனையொட்டி கணினிக் கலைச்சொல்லாக்கமும் சுருக்கமான சொற்களைக் கொண்டதாக அமைவது குறித்துப் பல்வேறு முயற்சிகள் மேற்கொள்ளப்படுகின்றன. சொல்லானது சுருக்கமாக அமையும் போது, பயன்பாட்டில் எளிமையானதாகவும், தெளிவுபடுத்துவனவாகவும் அமைகிறது.

Computer என்ற சொல்லானது தொடக்கத்தில் கணிப்பான் எனக் குறிப்பிடப் பெற்றது. ஆனால் Calculator என்னும் கருவியும் பயன்பாட்டில் இருந்ததால் 'கணிப்பான்' என்பதற்கு மாறாக 'கணிப்பொறி' என்னும் தொகைச்சொல் பயன்படுத்தப்பட்டது. ஆனால் கலைச்சொல்லாக்கத்தில் சிக்கனம் என்பது முக்கியமானதாக அமைகின்ற காரணத்தால் இன்று அது 'கணினி' என்னும் எளிய, தெளிந்த சொல்லாக வழங்கி வருகிறது.

இலக்கணச் சொற்களைக் கையாளல்

இலக்கணச் சொற்களைக் கணினிக் கலைச் சொற்களாகப் பயன்படுத்தும் முறையும் கையாளப்பட்டு வருகிறது. எழுவாய், பயனிலை, செயப்படுபொருள் என்பதில் வரும் பயனிலை என்பதை விரித்தால் பயன்நிலை என வரும். Application என்பதற்கும் பொதுவாக பயன்சார், பயன்பாடு ஆகிய சொற்களும், functional என்பதற்கு செயல்நிலை என்னும் சொல்லும் பயன்படுத்தப்படுகின்றன. ஆனால் verb என்னும் சொல் 'பயனிலை' என்று மொழிபெயர்க்கப்படுவதால் Application என்பதைப் பயனிலை என்று குறிப்பிடுவது பொருள் மயக்கத்தை உண்டாக்குவதால் அதனை 'பயன்' என்று மட்டுமே சுருக்கிக் கூறுவது சாலச் சிறந்தது. அவ்வாறே Functional என்பதை 'செயல்' எனக் குறிப்பிடுதலே போதுமானது.

நிறைவாக

கணினிக் கலைச்சொல்லாக்கத்தில் தமிழ் வழக்குக்கேற்ப இலக்கண விதிகளைப் பின்பற்றியே பெரும்பாலான புதுச்சொற்கள் படைக்கப்படுகின்றன. பண்டைக்கால வழக்கிழந்த சொற்களுக்குப் புதிய வடிவம் கொடுக்கப்பட்டு மக்களின் பயன்பாட்டில் இடம் பெற கலைச்சொல்லாக்கம் வழிவகுக்கின்றது. படித்தவர்களும், பாமரர்களும் கணினிப் பயன்பாட்டினை எளிமையாகப் பெறவும், கருத்தை எளிமையாகப் புரிந்துகொள்ளவும் புதுச் சொற்கள் அமைகின்றன. புதுப்புது சொல்லாக்கங்கள் தமிழ்மொழியில் உருவாக்கப்படும்போது செம்மொழித் தகுதிக்கு மேலும் வலுசேர்ப்பதாகவும், தமிழ்மொழியின் வளர்ச்சிக்குத் துணைபுரிவதாகவும் அமைகின்றன.

எனவே தமிழ்ச் சூழலுக்கு ஏற்ப, தமிழர்களின் பயன்பாட்டில் இடம் பெறும் வகையில் கணினிக் கலைச்சொல்லாக்கத்தில் புதுச்சொற்கள் படைக்கப்பட்டால் தான் கணினித்துறையில் தமிழ் இன்னும் கோலோச்ச முடியும் என்பது உறுதி.

துணைநின்ற நூல்கள்

1. இராதாசெல்லப்பன், கணினிக் கலைச்சொற்கள், நியூ செஞ்சுரி புக் ஹவுஸ் (பி) லிட், அம்பத்தூர், சென்னை-98, 2005.
2. இராதாசெல்லப்பன், கலைச் சொல்லியல், தாமரை பப்ளிகேஷன்ஸ் சிட்கோ இண்டஸ்டிரியல் எஸ்டேட், அம்பத்தூர், சென்னை-98, 2006.
3. சிதம்பரநாதச்செட்டியார், அ., புதுச்சொற்கள், கலைக்கதிர், பூ.சா.கோ. அறநிலையம், கோவை. 1957.
4. மாதையன், பெ., அகராதியியல், கலைச் சொல்லகராதி, பாவை பப்ளிகேஷன்ஸ், இராயப்பேட்டை, சென்னை-14, 2009.
5. ஜானகிராமன், நா., அறிவியல் கலைச் சொல்லாக்கம், இராகவன் பதிப்பகம், பெரியவடவாடி, 2003.

Transliteration Schemes for Tamil to Roman and Roman to Tamil characters

Dr.S.Srinivasan

Scientific Officer, Computer Division
Indira Gandhi Centre for Atomic Research
Kalpakkam-603102, Tamilnadu, India
srini@igcar.gov.in

Introduction

Here a machine transliteration scheme is proposed to map the random striking on a Tamil keyboard to Roman equivalent characters and vice versa. Many a language spoken in this world is associated with a native script. However a group of languages might also share a common script. For instance, the Roman script is shared by a number of European languages viz., English, French, German, Italian and Spanish. If there is a means to spell out a foreign language using ones native script it lessens the burden of mastering another script and speeds up the process of learning. In such a situation, the transliteration scheme comes handy and helps to overcome the difficulty of knowing yet another script.

The Tamil keyboard layout

The mechanical Tamil typewriter consists of 4-tier structure and encompasses 23 consonants, 12 vowels, about 12 vowel modifiers, about 20 ukara vowel-consonants, one medial and a conjunct (Sri). Besides these the keyboard also contains shift key, dead key, caps lock and space bar.

There are 4 schemes that are widely used for Tamil transliteration.

They are:

1. Madras University Tamil Lexicon Scheme- ISO 15919 Standard
(Based on lower ASCII and a few diacritical markers)
2. Library of Congress scheme (based on lower ASCII)
 - (a) It uses markers from upper ASCII block, i.e., characters 128-255.
 - (b) Variant of Library of Congress scheme (uses special characters such as # \$ _)
3. University of Koeln-Institute of Indology and Tamil Studies scheme (based on lower ASCII, case sensitive and also uses digit 2)
4. ITRANS developed by Avinash Chopde (uses special characters such as ~ ^ besides other roman characters)

The objectives of the transliteration scheme are as follows:

- 1.The transliteration scheme could be case sensitive but shall not employ any special character or digit.
- 2.The transliteration must be unambiguous (loss less) and is primarily meant for machine automation.
- 3.The transliteration may not be ideally suited for human reading. Hence its user-friendly aspect is kept secondary.

Tamil-to-Roman transliteration

Imagine an ape to snatch away a Tamil keyboard from a human hand and whimsically attempts to tap the keyboard to generate a sequence of characters. The strings of characters that are generated in the process may not be meaningful and some of them even contain a number of vowel modifiers occurring in succession (e.g. kombu,kaal,kombu...). To map such a sequence of modifiers, an improved rigid transliteration scheme is proposed. It is basically a variant of the UKoeln scheme.

None of the existing transliteration schemes are 100% romanized. All these schemes use either diacritical marker or special character or digit in addition to Roman letters. The UKoeln-IITS scheme uses the digit 2 to map the consonant Rannagaram and not any other special characters. Also this scheme does not differentiate between the pure vowel and its modifier. Hence attempt to improve upon this scheme was pursued.

The mapping for 'Rannagaram' was changed from n2 to nx so that the scheme becomes fully romanized. The choice for the letter x in lieu of 2 is the following.

nx => n(ExCHANGED) => 'tannagaram' exchanged to 'Rannagaram'

As the last bogie of a passenger train is indicated by the symbol X so also one can construe x to be the last consonant of archaic Tamil(ய்) and capital X to be the augmented last consonant of Grantha(ᱵ) to the Tamil character set.

The word குர்ஆன்(kurAn) was not properly transliterated from Tamil to Roman and then back to Tamil. Recasting the pure vowels from the vowel modifiers solved this problem. The vowel modifiers used in UKoeln-IITS scheme are:

a A i I u U e E ai o O au

To make the pure vowel representation unique the respective vowel modifiers were prefixed with capital Y and the result is,

Ya YA Yi YI Yu YU Ye YE Yai Yo YO Yau.

The choice of the letter Y springs from the fact that the pure-vowel sounds close to the yakara vowel-consonant(uyirmey).

e.g.

ஆறு => யாறு; ஆனை => யானை; ஆண்டு => யாண்டு; ஆர் => யார்;

எமன் => யமன்; எந்திரம் => யந்திரம்; (பரந்தாமனார், அ.கி. 1955)

உரேனியம் => யுரேனியம்; ஊகம் => யூகம்; Yiddish => இட்டிஷ்;

Yield => ஈல்ட்; Yes => எஸ்; Yellow => எல்லோ

The prefixing of the letter Y for pure vowels may not appear weird if one compares the following transliterated words:

e.g. கடை =>kaTai ; கதை =>katai

Also in Tamil the pure vowels occur at the beginning of words. They seldom occur in the middle of words (aLapeTai is the exception). On the contrary the yakara vowel-consonants seldom occur at the

beginning of words (yA is the exception). Case markers and words beginning with pure vowels when combine with a preceding Tamil word they change to yakara vowel-consonant.

e.g.

உடை/யணிந்து, இனிமை/யான, குடி/யிருப்பு, தலை/யீடு, உடை/யுடுத்து, உணர்ச்சி/யூட்டும், கை/யெழுத்து, இடை/யே, உதவி/யை, கை/யொப்பம், எத்தனை/யோ, நடக்கிறபடி/யால், மொழி/யில், அமைதி/யுடன், உதவி/யை, ஒலி/யோடு

Hence to differentiate between the pure vowel and its modifier the letter Y was chosen to be a part of the pure vowel. In English there are only two nasals (n and m). But in Tamil there are six nasals. Hence many of the nasals in Tamil need to be represented using more than one roman letter. Three of the nasal consonants (i ĩ ŷ) and a grantha consonant (ᳵ) have to be represented by two Roman letters. In such cases the second letter (trailing part) was chosen to be unique and did not figure in any of the other consonant representations. If any phonetic semblance required its alternate case was chosen. The scheme employs 14 upper case letters of which 5 are used to represent the long vowels, one as prefix to pure vowels, 7 for consonants and one for conjunct vowel-consonant.

A I U E O, Y, N T R L J H X, S

The proposed transliteration table is given below.

க் ச் ட் த் ப் ற்

k c T t p R

ய் ர் ல் வ் ழ் ள்

y r l v z L

ங் ஞ் ண் ந் ம் ன்

ng nj N n m nx

ஸ் ஷ் ஜ் ஹ் ச்ஷ்

s sh J H X

• ஸ்ரீ

q SRI

க கா கி கீ கு கூ கெ கே கை கொ கோ கௌ

ka kA ki kI ku kU ke kE kai ko kO kau

அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ

Ya YA Yi YI Yu YU Ye YE Yai Yo YO Yau

A few of the problematic words are transliterated herewith.

e.g.

ஐந்து => Yaintu

அஇஅதிமுக => YaYiYatimuka (a popular political party in Tamilnadu)

குர்ஆன் => kurYAnx (from kriya Tamil dictionary)

குரான் => kurAnx

ஆகா ! => YAkA !

ஆகாஐ ! => YAkAAA

(one finds such usage in modern Tamil short stories- akin to emoticon usage in SMS and emails)

கை => kai

கைலாசம் => kailAcam

கஇலாசம் => kaYilAcam [வாழைப்பள்ளிக் கல்வெட்டு: கி.பி. 9-ஆம் நூற்றாண்டு]
(சிவலிங்கனார் அ. 1981)

தூது =>tUtu

துது =>tuAtu (thinamalar Tamil newspaper uses this way)

|கெ =>eke [archaic Tamil usage]: |+| =>'' (சம்பந்தன், மா.சு. 1981)

Also the split vowel modifier of aukaaram(ள) is construed as a combination of the modifiers kombu(!)
and kaal(j): |j| =>eA; |+j| => ள

Evidence to support this contemplation exists. (சிவசேகரம் சி. 1993)

The refinement of Tamil characters is also going on at slow pace. Reverend Veeramaa Munivar for the first time in modern era brought out orthographic refinement in Tamil. He devised a method to remove the ambiguity in the appearance of long and short vowels of ekaram and okaram about 250 years ago. The next refinement of Tamil characters initiated by the Tamilnadu Government and coinciding with the birth centenary year of Periyaar (E.V.Ramasamy) took place in the year 1979. To refine pure vowels with minimal change R.Krishnamurthy, the editor of Thinamalar Newspaper advocates the following modification for long vowels(கிருஷ்ணமூர்த்தி, இரா, 1978).

அ அி இ இி உ உி எ எி ஒ ஒி

Ya YaA Yi YiA Yu YuA Ye YeA Yo YoA

For instance Thinamalar newspaper recasts the Ukaarams œ, ß, à, á, æ, ê, ë to šj, İj, Đj, Ñj, Öj, Új, Ūj and implements them in its edition. The aforesaid roman transliteration scheme is robust enough to take care of any foreseeable refinements in Tamil characters as well. The following is a sample text taken from Bharathidasan's work and whose transliteration is also given alongside.

அழுபவன் கோழை	Yazupavanx kOzai
ஆவின் பால் இனிது	YAVinx pAl Yinxitu
இரவினில் தூங்கு	Yiravinxil tUngku
ஈவது மகிழ்ச்சி	YIvatu makizcci
உள்ளதைப் பேசு	YuLLataip pEcu
ஊமைப்போல் இராதே	YUmaippOl YirAtE
எதையும் ஊன்றிப் பார்	Yetaiyum YUnxRip pAr
ஏசேல் எவரையும்	YEcEl Yevaraiyum
ஐந்திற் கலை பயில்	YaintiR kalai payil
ஒற்றுமை வெல்லும்	YoRRumai vellum
ஓரம்போ தெருவில்	YOrampO teruvil
ஒளவை தமிழ்த்தாய்	Yauvai tamizttAy
கணக்கில் தேர்ச்சிகொள்	kaNakkil tErccikol
சரியாய் எழுது	cariyAy Yezutu
தமிழ் உன் தாய்மொழி	tamiz Yun tAymozi

Roman-to-Tamil transliteration

A rigid transliteration scheme is contemplated for Roman-to-Tamil conversion for the benefit of Tamils who can read only Tamil characters but wish to read English text as well. A parallel exists to this. Historically Tamils invented an alternate script to read Sanskrit text. It was called the Grantha. The Tamils preferred this script in lieu of the Devanagari script as it contained the Tamil alphabet too. In the same lines, an attempt has been made to map both the upper and lower case Roman letters (English characters) into Tamil.

Let us perform another thought type experiment. Imagine another ape belonging to North-Atlantic region to grab away a Roman keyboard from a human hand and strikes whimsically on the keyboard. Also assume that the digit and special character keys are disabled from the keyboard. In such a situation too, the string of characters that are generated may not appear meaningful and could even contain a large mix-up of both upper and lower case letters as nouns appear in German text. To map such a sequence of letters into Tamil a transliteration scheme is proposed. The contemplated case inclusive transliteration table is given below.

lower case								
------------	--	--	--	--	--	--	--	--

a	b	c	d	e	f	g	h	i
அ	•ப	ச	ட	ஏ	•வ	•க	ஹ	இ
j	k	l	m	n	o	p	q	r
ஐ	க	ல	ம	ந	ஒ	ப	கு	ர
s	t	u	v	w	x	y	z	
ஸ	த	உ	வ	வு	சஷ	ய	•ஜ	

UPPER CASE								
------------	--	--	--	--	--	--	--	--

A	B	C	D	E	F	G	H	I
ஆ	•ப	ச	ட	ஏ	•வ	•க	ஹ	ஈ
J	K	L	M	N	O	P	Q	R
ஐ	க்	ல்	ம்	ந்	ஓ	ப்	க்	ர்
S	T	U	V	W	X	Y	Z	
ஸ்	த்	ஊ	வ்	வூ	சஷ	ய்	•ஜ்	

With the aid of the above table the following English words are transliterated into Tamil.

apple-	அபபலஏ	APPLE-	ஆப்ப்ல்ஏ
box-	•பஒசஷ	BOX-	•ப்ஒசஷ
cat-	சஅத	CAT-	ச்ஆத்
dog-	டஒ•க	DOG-	ட்ஒ•க்

elephant-	எலஎபஹஅநத	ELEPHANT-	எல்எப்ஹஆந்த்
fox-	•வஓசஷ	FOX-	•வஓசஷ
goat-	•கஓஅத	GOAT-	•க்ஓஆத்
horse-	ஹஓரஸஎ	HORSE-	ஹ்ஓர்ஸ்எ
ink-	இநக	INK-	ஈந்க்
jug-	ஜஉ•க	JUG-	ஜ்ஊ•க்
kite-	கஇதஎ	KITE-	க்ஈத்எ
lilly-	லஇலலய	LILLY-	ல்ஈல்ல்ய்
man-	மஅந	MAN-	ம்ஆந்
nose-	நஓஸஎ	NOSE-	ந்ஓஸ்எ
owl-	ஓவுல	OWL-	ஓவுல்
pig-	பஇ•க	PIG-	ப்ஈ•க்
quill-	குஉஇலல	QUILL-	க்ஊஈஈல்
rat-	ரஅத	RAT-	ர்ஆத்
snake-	ஸநஅகஎ	SNAKE-	ஸ்ந்ஆக்எ
turkey-	தஉரகய	TURKEY-	த்ஊர்க்யய்
umbrella-	உம•பரஎலலஅ	UMBRELLA-	ஊம்•ப்ர்எல்ல்ஆ
van-	வஅந	VAN-	வ்ஆந்
window-	வுஇநடஓவு	WINDOW-	வுஈந்ட்ஓவு
xmas-	சஷமஅஸ	XMAS-	சஷம்ஆஸ்
yacht-	யஅசஹத	YACHT-	ய்ஆச்ஹத்
zero-	•ஜஎரஓ	ZERO-	•ஜ்எர்ஓ

The reading of this transliteration may seem difficult in the initial phase. This situation is akin to asking a German to read English text or an English man to read German text. But this difficulty can be overcome in due course with adequate practice. The poet Bharathidasan himself felt that the alphabetical system of Tamil could be simplified if all the vowel consonants were split into pure vowels and consonants (பாரதிதாசன், 1948).

மறைவாக	நமக்குள்ளே	பழங்கதைகள்
(ம்அற்றஐவ்ஆக்அ	ந்அம்அக்கஉள்ளஎ	ப்அழ்அங்கஅத்ஐக்அள்)

சொல்வதிலோர் மகிமை இல்லை
(ச்ஓல்வ்அத்இல்ஓர் ம்அக்இம்ஐ இல்லஐ)

He tried even writing so; but discontinued this effort in due course.

In a nutshell the motive behind this scheme is the following:

1. To transliterate the lower case Roman letters into Tamil the following Tamil letters are used. They are short vowels, consonants, medial and a few ukara vowel-consonants.
2. To generate the mapping for the upper case Roman letters from the lower case Roman letters is simple. They consist of either long vowels or pure consonants.

Applications

A random sequence of characters that are generated from a keyboard can be faithfully transliterated to another script and can also be used as a substitute for a password in computer applications. A password of this kind is all the more difficult to crack by a hacker. Transliterating a text to a non-native script and then encrypting it would add to the level of data security. The mapping from Tamil-to-Roman involves variable length and dual case and hence once encrypted it is difficult to decrypt. This feature enhances the data security in transmission as well as storage.

Conclusion

A machine transliteration scheme is proposed to map all the random sequence of characters that could be generated from a Tamil or a Roman keyboard. The transliterated characters could further be encrypted so as to increase the level of security.

Acknowledgements

The author expresses thanks to Dr.Vasu Renganathan, University of Pennsylvania, Philadelphia, USA and Dr.K.Kalyanasundaram, Lausanne, Switzerland for their help in providing the various romanized transliteration standards available for Tamil and offering valuable comments on this work.

References

1. URL: <http://homepage.ntlworld.com/stone-catend/translit.htm>
2. URL: <http://www.aczone.com/itrans/tamil/node5.html>
3. பரந்தாமனார், அ.கி. 1955, நல்ல தமிழ் எழுத வேண்டுமா? சென்னை: பாரி நிலையம்.
4. சிவலிங்கனார் அ. 1981, தொல்காப்பியம் -எழுத்ததிகாரம் மொழி மரபு, சென்னை: உலகத் தமிழாராய்ச்சி நிறுவனம்.
5. சம்பந்தன், மா.சு. 1981, எழுத்தும் அச்சும், சென்னை: தமிழர் பதிப்பகம்
6. சிவசேகரம் சி. 1993, தமிழும் அயலும், தேசிய கலை இலக்கியப் பேரவை, சென்னை: சூர்யா அச்சகம்.
7. கிருஷ்ணமூர்த்தி, இரா. 1978, தமிழ் எழுத்துச் சீர்திருத்தம். சென்னை: தினமலர் வெளியீடு.
8. Krishnamurthy, R. 1977, Script Reform in Tamil, Seminar on Socio Linguistics and Dialectology, March 27, Annamalai nagar: Annamalai University.
9. பாரதிதாசன், 1948, மே 15, குயில் திங்களிதழ், புதுவை.

From Classical Tamil to Computational Tamil A Perspective

கன்னித்தமிழிருந்து கணினித்தமிழுக்கு: ஒரு கண்ணோட்டம்

Dr A Kumaran
Microsoft Research India
Bangalore, India.

Abstract

The area of Computational Linguistics deals with computational models that are employed for analysis, synthesis or transformation of content in natural languages. Many well known end user technologies, such as, language understanding, machine translation, monolingual and crosslingual information retrieval and extraction, etc., are based on such models. Given the exponential growth of content in the Internet and Social Media primarily in vast majority of languages of the world, it is highly imperative that tools and technologies be developed to process the natural language data effectively and efficiently. In this paper, we highlight the state-of-the-art approaches for Computational Linguistics that are primarily based on statistical and machine learning principles, and underscore the need for clean large annotated corpora and language resources for any and all types of Computational Linguistics research. In particular, we emphasize the need for corpora, basic tools and resources in Tamil, in order to ensure the development of technologies in the Tamil language. It is imperative that the community, academia, industry and the government come together to create a climate of consensus, coordination and collaboration to make sure that Tamil is taken successfully to the computational world.

Introduction & Motivation

The area of Computational Linguistics deals with computational models for analysis and synthesis of natural languages, and is a vital predecessor for many natural language processing tasks, such as language understanding, summarization, information retrieval and extraction, machine translation, etc. Given the exponential growth of the amount of available natural languages data due to the Internet and Social Media, it is highly imperative that tools and technologies be developed to process the data effectively and efficiently. More importantly, in countries like India where only about 5% of the people are English literate, the need for such technologies is even more important for including the majority into the Information Age.

Computational Linguistics research pertains to development of such tools and technologies. Traditionally, Computational Linguistics research and systems relied on linguistics research resulting in rules that are distilled by experts in a language. For examples, rules that govern morphological variations of a word or formation of a sentence in a given language are devised by experts, and coded into practical tools and systems. However, given that the natural languages evolve, such systems become unmanageable as they are fraught with relatively large portion of exceptions for every rule. Further, such rule-based

approaches are expensive to create and maintain in terms of time and resources, as evidenced by decades of research put in the Western European (WE) and Chinese-Japanese-Korean (CJK) languages.

In the recent decade and a half, a host of newer approaches has been introduced in the Computational Linguistics Research, specifically, those based on Statistical and Machine Learning based methodologies. In these methodologies, specific tasks may be *learnt* automatically when provided with appropriate hand-crafted training data. These methodologies are broadly referred to as *statistical learning* or *machine learning* algorithms. For example, identification of names or places in a sentence may be learnt (with a certain level of accuracy) by programs that are trained on large hand-annotated corpora, so that they may be used subsequently for identification of names from sentences. While the quality of such depends on several factors: such as, the nature of the task, the algorithms used, features used for training, the quality and quantity of data used for training, etc., still such approaches had been proven to be very effective – as good or better than hand crafted systems for many of the natural language processing tasks. For example, all the state-of-the-art-translation systems in the world now are statistical learning systems.

In addition to being easier to develop, equally importantly, such methodologies are also largely language-independent, paving way for quick adaptation across languages. For example, a generic Statistical Machine Translation (SMT) system may be employed successfully to learn translations between any given pair of languages (with appropriate training data in those pair). Hence, such approaches exhibit a great advantage especially in countries like India, where a single system may be adopted for many languages, quickly and transparently. In essence, these methodologies rely on generic statistical and machine learning frameworks, trained on custom datasets.

Finally, given that the most popular medium for information, entertainment, commerce and governance – The Internet – is also turning multilingual³¹. The demographics of Internet users have changed from being predominantly English, to more than two-thirds that are non-native English speakers now. In addition, majority of the information available over the web is in a language other than English. Such shifts in demographics suggest that the technologies must be developed for supporting predominantly multilingual user population, pointing to the critical need for language neutral Computational Linguistics research to cater to wider audience, quicker. In countries such as India, we face additional challenges where the population is mostly English illiterate³², hence tools and technologies in local languages are even more important, in order to overcome the digital divide to include the common man.

In the subsequent sections, we specify the type of corpora needed for Computational Linguistics research, and appeal to the Tamil linguistics and computational linguistics community to work toward common sets of standards and corpora to make the research community vibrant and fruitful.

Linguistic Corpora to be Developed

In this section, we outline several types of linguistic standards and corpora that need to be developed for Tamil, to support robust computational linguistics research.

³¹ <http://www.GlobalReach.biz>.

³² It is estimated that only about 5-7% of the Indian population is conversant in English.

National Efforts on Linguistic Corpora

National corpora are normally general reference corpora which are supposed to represent the national language of a country. They are collected by a concerted effort by the Government along with Academic and/or Industry players, in a focussed manner. These corpora are balanced with regard to genres and domains that typically represent the language under consideration, in that particular geographic or political domain. While many of the national corpora are available with parts of speech annotation, few of them have syntactic and semantic parses annotated.

The British National Corpus (BNC) is perhaps the first and best-known national corpus. It is designed to represent as wide a range of modern British English. This comprises approximately 100 million words of written texts (90%) and transcripts of speech (10%) in modern British English. In addition to part-of-speech (POS) information, the BNC is annotated with rich metadata (i.e. contextual information). The American National Corpus (ANC) project was initiated in 1998 with the aim of building a corpus comparable to the BNC. The first release of the corpus contains 11.5 million words of written and spoken data. When completed ANC will contain a corpus of (100M) words comparable to BNC corpus. The corpus is POS tagged using different tag-sets to suit the needs of different users. Similarly there are national corpora available in Polish (130.8M), Czech (100M), Russian (100M), Hellenic (32M), German (100M), and Chinese (700M characters) languages.

In India, Central Institute of Indian Languages (CIIL) collected corpus is available in most Indian languages. However, this corpus is a relatively small corpus (approximately 3-8M words per language) primarily a monolingual text collection in multiple languages, with no annotation. While this corpus may provide the seed for data creation, the volume and quality of such corpus needs to be enhanced significantly, to aid Computational Linguistics research in Indian languages.

Recently, Linguistic Data Consortium for Indian Languages (LDC-IL) has been initiated by the Ministry of Human Resource and Development under Government of India, to oversee the standardized collection of linguistic corpora in all Indian languages. Several academic and industrial partners are working together to get this collection created.

Monolingual Corpora

Monolingual corpora essentially refer to normal Tamil language text bulk from a standard source, such as, popular mass media, newspapers, television, etc. While it is good to have a wide variety of content; each genre, say, printed or spoken news, literary works, political speeches, religious writing, etc., each has its own characteristics, and best handled individually.

Ideally the text bulk should be in a standard encoding, such as Unicode, and annotated with some meta-data, such as, source, author, date of publication, genre or category, etc. In addition to document level annotation, annotation of the content of the corpus itself could be extremely useful for many Computational Linguistics Tasks. For example, a corpus annotated with names (personal names, common names, places, dates, organizations, etc.), may be used for Named Entity identification tasks, and Information Extraction tasks.

Multilingual Corpora

Multilingual corpora refer to many types of corpora – parallel, comparable, etc. Parallel corpora are essentially sentence aligned corpus in multiple languages, where every aligned sentence pair contains the same semantic information in multiple languages. Such corpora may be readily used for developing Machine Translation systems.

In many practical situations, comparable corpora are more readily available than parallel corpora. Comparable corpora are defined as article aligned corpus in multiple languages, where the article generally is on the same topic, but may have different semantic content. Typical comparable corpora consist of news articles in multiple languages that cover the same news event; since each article may be written by different editor, it is likely to have similar, but not the same semantic content. Comparable corpora had been successfully employed in development of MT systems.

Annotated Corpora

Large annotated corpora are critically needed in any computational linguistics research. The annotation depends on the task at hand; for example, Part of Speech (POS) identification requires a rich annotation where every word in the text corpus is tagged, whereas Named Entity Recognition (NER) requires hand annotation of specific entities in the corpus.

Annotation Standards

Any type of annotated corpora underscores the need for annotation standards, in order to create standard annotated corpora that may be used by many tools and research groups consistently. It is imperative that standards be developed for annotation of the collected data such that the data created is rich enough to support many learning tasks that need to be based on the data, yet, flexible enough to be modified when the need arises. An example of such effort, is given in (Baskaran et al., 2008), where a Part of Speech annotation framework – called IL-POST – was designed collaboratively by a set of academic and industry partners and which is applicable to a variety of Indian languages. Standard frameworks exist, such as EAGLE's (Leech et al., 1996) for linguistic tagging.

Linguistic Resources

Many resources, such as dictionaries, thesauri, Bilingual or multilingual dictionaries are necessary for obvious reasons, for a variety of tasks. A computational dictionary must be in a standard format (Unicode, XML tagged) and must be machine-readable with standard tags. In addition, all references linking various words (for example, in thesauri) must be navigable using unique identifiers. There are specific requirements for computational dictionaries, as against print dictionaries.

Standards Organizations

Over the last few decades there are many very successful initiatives between governments, industry and academia in developing standards and corpora according to those standards. An example of such an initiative is the Linguistic Data Consortium (LDC) in the University of Pennsylvania.

Creation of Data with Community-wide Participation

It is important to highlight the importance of crowd-sourcing as a methodology for creation of linguistic corpora, as many types of corpora does not need to be created by linguists or language experts, but easily

by the native speakers of a language. An initiative to generate parallel data is outlined in (Kumaran et al., 2009).

The Need for Linguistic Corpora

In this paper, we focused on mining NE pairs in two different languages, namely English and an Indian language, Tamil. While we adopted a methodology similar to that in [Klementiev and Roth, 2006], our focus was on mining parallel NE transliteration pairs, leveraging the availability of comparable corpora and a well-trained linear classifier to identify transliteration pairs. We profiled the performance of our mining framework on several parameters, and presented the results. While the results show the potential of our approach, we also uncovered several issues that need to be resolved, for effective mining of parallel NE transliteration pairs. Given that the NE pairs are an important resource for several NLP tasks, we hope that such a methodology to mine the comparable corpora may be fruitful, as comparable corpora may be available in perpetuity in several of the world's languages.

References

1. Baskaran, S., Bali, K., Bhattacharya, T., Bhattacharyya, P., Choudhury, M., Jha, G. N., Rajendran, S. 5., Saravanan, K., Sobha, L., and Subbarao, K. V. S. 2008.. **A Common Parts-of-Speech Tagset Framework for Indian Languages..** In *Proceedings of LREC 2008*, Morocco.
2. Kumaran, A., Saravanan, K., Datha, N., Ashok, B. and Dendi, V. 2009. **WikiBABEL: A wiki-style platform for creation of parallel data.** In *Proceedings of ACL 2009*.
3. Leech, G and Wilson, A. 1996. **Recommendations for the Morphosyntactic Annotation of Corpora.** EAGLES Report EAG-TCWG-MAC/R.
4. Linguistic Data Consortium. <http://ldc.upenn.edu/>.

Automated Identification of Grammatical Patterns For Tamil Poems

Sendhilkumar S., Mahalakshmi G.S., Prakash N

Department of Computer Science and Engineering,
Anna University, Chennai 600025,
thamaraikumar@cs.annauniv.edu,
mahalakshmi@cs.annauniv.edu,
yoursprakashever@gmail.com

Abstract

Tamil is one of the most classical world languages. But in modern teaching, people have started ignoring it. One main reason is that the grammar is comparatively tougher than that of other languages. The aim of this prototype is to build an application which makes teaching and learning Tamil grammar more easily. With rapid advancement in Tamil language processing packages, it is not very tough to build an interactive, user-friendly tool which will help in teaching-learning Tamil grammar in a better manner. The idea is to develop rule-base for calculating மாத்திரை and later, identification of grammatical patterns present in the Tamil poems as well as classification of Tamil poems based on the structure becomes more flexible.

Introduction

In recent days, a survey on natural language processing tools for Tamil results in a handful of research approaches and solutions, of which a few are officially recognized. There is enough research done to analyze the morphemes of a word or sentence (say, morphological analyzer, morphological generator, POS tagger, parser etc.), but usage of such NLP packages are yet to touch the real needy. The need for encouraging people to learn Tamil lies truly in the application of information and communication techniques to teach basic grammar in school curriculum, which is now a problem not yet addressed. This paper will be an initiative which will try to help school children a lot by encouraging them to use ICT for learning Tamil grammar in a more interesting manner.

Methodology

The three of five major divisions of Tamil grammar say, எழுத்திலக்கணம், யாப்பிலக்கணம் and சொல்லிலக்கணம் are considered for developing the application prototype. Rule base is constructed for calculating மாத்திரை. The occurrences of குற்றியலுகரம், குற்றியலிகரம், மகரக்குறுக்கம், உயிரளபெடை, முற்றியலுகரம், ஒற்றளபெடை, ஆய்தக்குறுக்கம், ஐகாரக்குறுக்கம் and ஒளகாரக்குறுக்கம் are identified and rules are generated. In addition, the classification of poems (வெண்பா, ஆசிரியப்பா, அகவற்பா, கலிப்பா, வஞ்சிப்பா) based on their structure is done. Adding to that, rule base is created to find the presence of various grammatical items including எழுத்து, அசை, சீர், தளை, அடி, மோனை, எதுகை and பா வகை.

The existing morphological analyzers [2] address சொல்லிலக்கணம் in a decent manner. Utilizing the morphological analyzer, in this prototype, we shall analyze the words in the given input. For our application, we can get either TAB or Unicode input with the help of NHM writer. For common processing of both inputs, the input string is converted into internal codes based on a coding system (shown in Table 1) which just contains codes for அ,ஆ,இ,...,ஒள and க்,ஞ்,...,ன் and special

characters ஸ்,ஷ்,ஹ,ஜ்,சுஷ்,ஸ்ரீ. We do all the processing with these internal codes and while displaying the output, the results are converted back into corresponding TAB/Unicode encoding.

For example, the input string அம்மாஅப்பா it will be converted into the byte array which contains the internal codes 1 23 23 2 1 22 22 2.

Table 1. Internal Encoding

	0	10	20	30
0		ஒ	த்	ற்
1	அ	ஓ	ந்	ன்
2	ஆ	ஒள	ப்	ஸ்ரீ
3	இ	ஃ	ம்	ஷ்
4	ஈ	க்	ய்	ஸ்
5	உ	ங்	ர்	ஹ்
6	ஊ	ச்	ல்	சுஷ்
7	எ	ஞ்	வ்	
8	ஏ	ட்	ழ்	
9	ஐ	ண்	ள்	

எழுத்திலக்கணம்

1. மாத்திரை

மாத்திரை is the time that should be taken to pronounce a word. குறில் will have 1 மாத்திரை, நெடில் will have 2 மாத்திரை, ஆய்தம் and மெய் will have 1/2 மாத்திரை. To calculate மாத்திரை, we should check the byte array we have got one by one. For உயிர்க்குறில் and உயிர்நெடில், we can assign மாத்திரை directly. In case of மெய், உயிர்மெய்க்குறில் and உயிர்மெய்நெடில், we should check for the next byte in the array before assigning மாத்திரை, because the next byte only will decide whether the character is a குறில் or நெடில் or மெய். If the next byte is உயிர்க்குறில், the character is உயிர்மெய்க்குறில். If the next byte is உயிர்நெடில், the character is உயிர்மெய்நெடில். If the next byte is not an உயிர், the character is a மெய்.

2. மாத்திரை reduction

Due to the presence of some special characters like குற்றியலுகரம், குற்றியலிகரம், மகரக்குறுக்கம், உயிரளபெடை, முற்றியலுகரம், ஒற்றளபெடை, ஆய்தக்குறுக்கம், ஐகாரக்குறுக்கம் and ஒளகாரக்குறுக்கம் there will be change (reduction) in the மாத்திரை of the character in which it is present. As a result, the total மாத்திரை of the word will also change. For குற்றியலுகரம், குற்றியலிகரம், ஐகாரக்குறுக்கம் and ஒளகாரக்குறுக்கம், the மாத்திரை of the character will get reduced by 1/2. For மகரக்குறுக்கம் and ஆய்தக்குறுக்கம், the மாத்திரை of the character will get reduced by 1/4. For உயிரளபெடை and ஒற்றளபெடை, there will not be any மாத்திரை change. The rules for identifying the presence of these special characters in a word are mainly obtained from [10]. The other books referred are [6],[7],[8],[9]. The

rules obtained are programmed in java and put into rule base. Applying these rules, the மாத்திரை reduction can be calculated.

We can give direct Tamil prose input or we can give already typed Tamil prose text file as input. In எழுத்திலக்கணம், if we click மாத்திரை கணக்கிடு button, the above output will be obtained. It will display the மாத்திரை of each character in a word and reduction of total மாத்திரை of the word due to the presence of special characters like குற்றியலுகரம், குற்றியலிகரம்,..... The sample output obtained for எழுத்திலக்கணம் is shown in Figure 1.

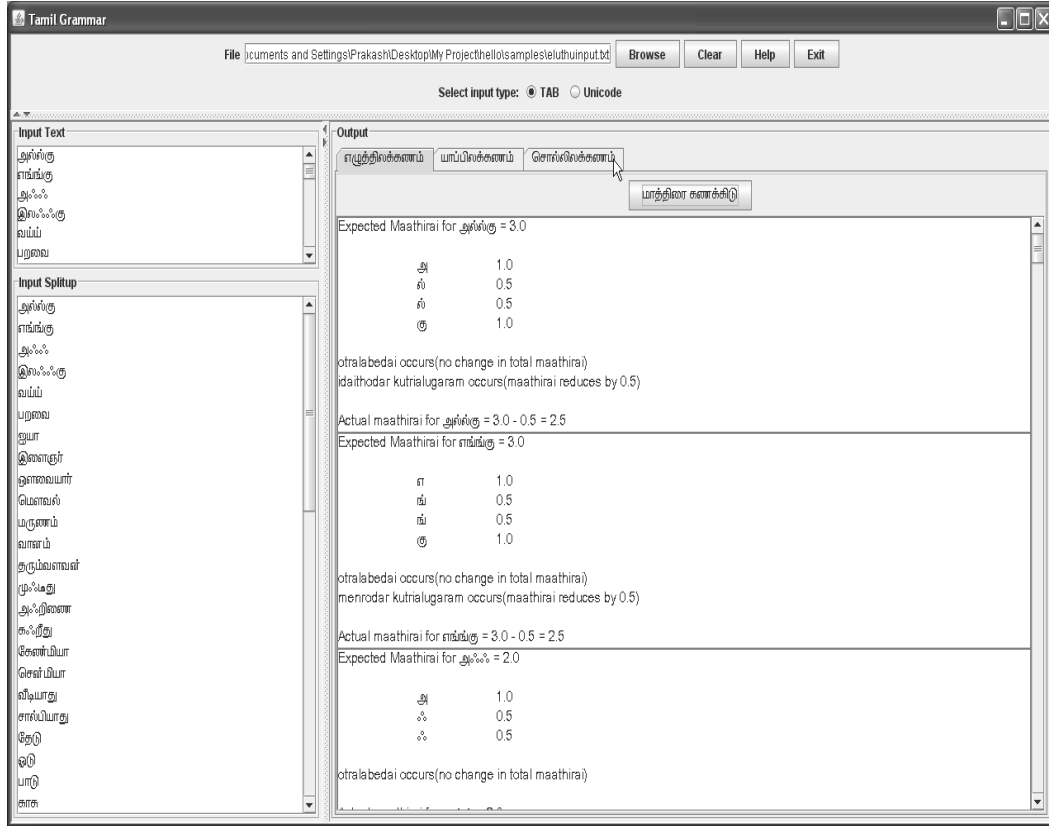


Figure 1. எழுத்திலக்கணம்

யாப்பிலக்கணம்

எழுத்து refers to each individual character in the given input. The character may be உயிர்க்குறில், உயிர்நெடில், மெய், ஆய்தம், உயிர்மெய்க்குறில் or உயிர்மெய்நெடில். Individual character or a group of characters together constitute an அசை. The two types of அசை are நேர் அசை and நிரை அசை. One or more அசை are grouped together to form a சீர். If one அசை forms a சீர், it is called ஓரசைச்சீர்(நாள், மலர், காசு, பிறப்பு). If two அசைகள் form a சீர், it is called ஈரசைச்சீர் or இயற்சீர்(தேமா, புளிமா,...). If three அசைகள் form a சீர், it is called மூவசைச்சீர். Two types of மூவசைச்சீர் are காய்ச்சீர் (தேமாங்காய், புளிமாங்காய்,...) and கனிச்சீர் (தேமாங்கனி, புளிமாங்கனி,..). Similarly four அசை also can form a சீர்(நாலசைச்சீர்). By checking the சீர் of the two consecutive words, the தளை between can be identified. The seven types of தளை are நேரொன்றாசிரியத்தளை, நிரையொன்றாசிரியத்தளை, இயற்சீர் வெண்டளை, வெண்சீர் வெண்டளை, கலித்தளை, ஒன்றிய வஞ்சித்தளை, ஒன்றாத வஞ்சித்தளை. Two or more சீர்கள்

together form an அடி (line of a poem). The types of அடி are குறளடி(2 சீர்கள்), சிந்தடி(3 சீர்கள்), அளவடி(4 சீர்கள்), நெடிலடி(5 சீர்கள்), கழிநெடிலடி(more than 5 சீர்கள்).

To make the poem sound good and meaningful, the characters and seers are arranged with care. One such attempt is named தொடை. The two main types are மோனை and எதுகை. மோனை means the first character of the two words which are compared are same. The two types of மோனை are சீர் மோனை and அடி மோனை. In சீர் மோனை, we compare the சீர்கள் in an அடி. There are 7 types of சீர் மோனை based on which சீர்கள் in an அடி we are comparing. In அடி மோனை, we compare the first சீர் of two lines (அடி).

எதுகை means the second character of the two words which are compared are same, provided the மாத்திரை of their first characters are also equal. The two types of எதுகை are சீர் எதுகை and அடி எதுகை. In சீர் எதுகை, we compare the சீர்கள் in an அடி. There are 7 types of சீர் எதுகை based on which சீர்கள் in அடி we are comparing. In அடி எதுகை, we compare the first சீர் of two lines (அடி).

Finally, we can find the பாவகை of the input poem with the number of different types of அசைகள் and சீர்கள் they are having and some special conditions. If the poem belongs to வெண்பா, ஆசிரியப்பா, கலிப்பா, வஞ்சிப்பா or அகவற்பா, we can identify it. We can give direct Tamil poem input or we can give already typed Tamil poem text file as input. In யாப்பிலக்கணம் tab if we press எழுத்து button, it will display the type of each character in a word like உயிர்க்குறில், உயிர்நெடில், மெய், உயிர்மெய்க்குறில், உயிர்மெய்நெடில், ஆய்தம். If we press அசை button, it will split up of நேர் and நிரை அசை in it. If we press சீர் button, it will categorize சீர் (மா, விளம், காய், கனி) based on அசைகள் they have. If we press அடி, it will categorize அடி into குறளடி, சிந்தடி, அளவடி, நெடிலடி, கழிநெடிலடி. If we press மோனை or எதுகை button, it will show if any மோனை or எதுகை present in it. If we press பாவகை button, it will show whether the poem belongs to வெண்பா, ஆசிரியப்பா, கலிப்பா, வஞ்சிப்பா based on சீர்கள் and தளைகள் they have. The sample output obtained for யாப்பிலக்கணம் is shown in Figure 2.

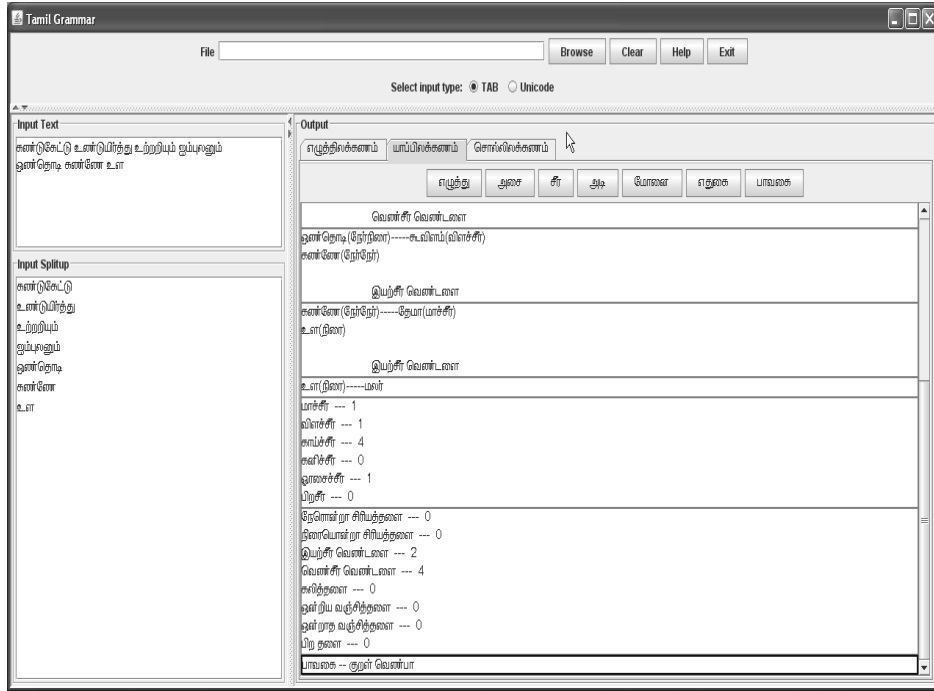


Figure 2. யாப்பிலக்கணம்

2.3 சொல்லிலக்கணம்

Tamil Morphological analyzer-atcharam [3], [4], [5] is used to get the output for சொல்லிலக்கணம். We can give direct Tamil prose input or we can give already typed Tamil prose text file as input. Various grammatical patterns like noun, verb, adjective, adverb, participle, clitics, singular, plural, accusative case, locative case, associative case, sandhi, infinitive suffix, permissive suffix, postposition, present tense maker, past tense maker, future tense maker, etc., can be identified by using this. The sample output obtained for சொல்லிலக்கணம் is shown in Figure 3.

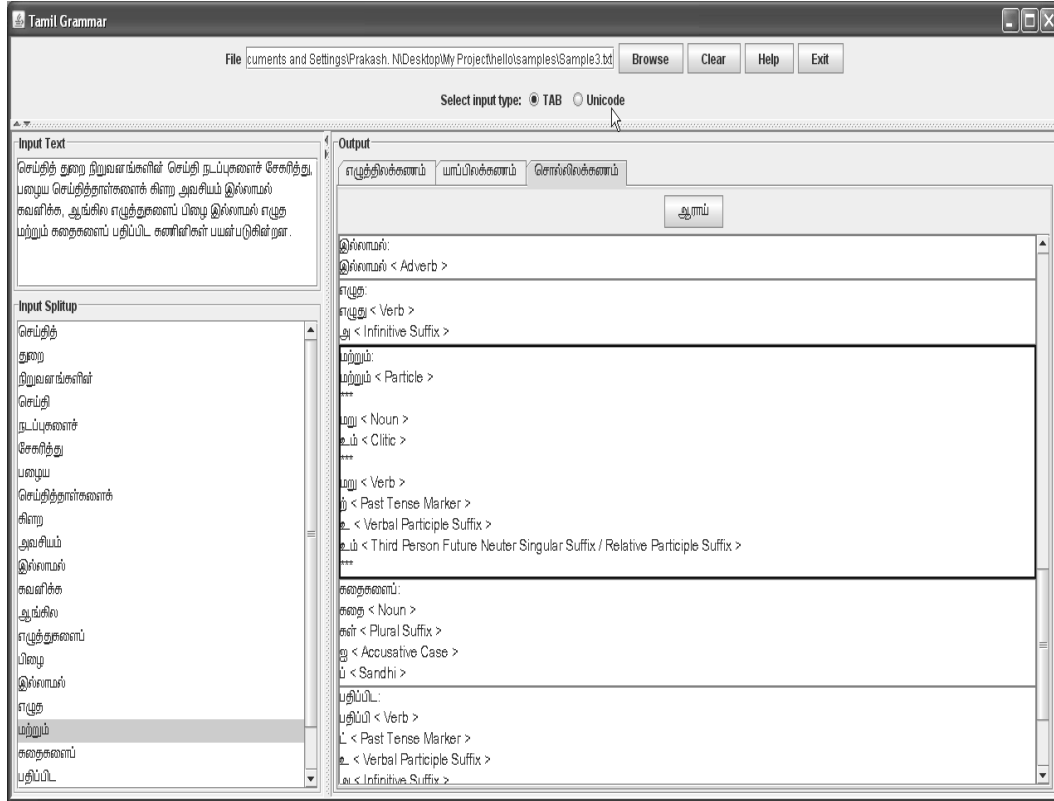


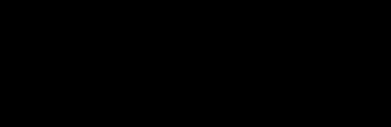
Figure 3. சொல்லிலக்கணம்

Conclusion

While utmost care has been taken to develop the grammar prototype in Unicode, it is no doubt that automatic identification of grammatical patterns is going to be a worthy application with respect to school education. பொருளிலக்கணம் and அணியிலக்கணம் deal with the semantics of the language. With decent Tamil dictionaries like agaradhi [1], handling பொருளிலக்கணம் is not an issue. However, அணியிலக்கணம் requires enough research thought to be thrown before things could be materialized. But, times are not too far that in future, using Tamil NLP tools, one should be able to analyse the similes present in the input Tamil text.

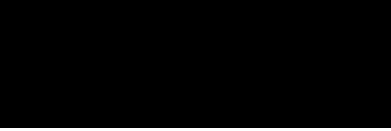
References

1. <http://www.agarathi.com/index.php>
2. <http://tdil.mit.gov.in/download/rctools/Atcharam.htm>
3. A.G. Menon, S. Saravanan, R. Loganathan, Dr. K. Soman, "Amrita morph analyzer and generator for Tamil: A rule based approach", in the proceedings of Tamil Internet Conference 2009, Koeln, Germany
4. M. Anand Kumar, V. Dhanalakshmi, S. Rajendran, K.P. Soman, "A novel approach to Morphological Analysis for Tamil Language", in the proceedings of Tamil Internet Conference 2009, Koeln, Germany
5. P. Anandan, Dr. Ranjani Parthasarathy & Dr. T.V. Geetha, "Morphological generator for Tamil", in the proceedings of Tamil Internet Conference 2001, KualaLampur, Malaysia,
6. செ.வை. சண்முகம், "எழுத்திலக்கணக் கோட்பாடு", உலகத்தமிழ் ஆராய்ச்சி நிறுவனம் வெளியீடு, 2001
7. தொல்காப்பியர், "தொல்காப்பியம்", downloaded from <http://www.projectmadurai.org/pmworks.html>, accessed latest by April, 2010
8. ந. கடிகாசலம், ச.சிவகாமி, "தொல்காப்பியப் பாவியல் கோட்பாடுகள்", உலகத் தமிழ் ஆராய்ச்சி நிறுவனம் வெளியீடு, 1999
9. பவணந்தி முனிவர், "நன்னூல்", downloaded from <http://www.projectmadurai.org/pmworks.html>, accessed latest by April, 2010
10. புலவர் ஏ. லோகநாதன், "தமிழ் இலக்கணத் தெளிவுரை", இளமதி பதிப்பகம், February 2006



4

**கணினி வழி தமிழ் மொழி
சொந்திருத்திகள்**



தானியங்கு சொற்பிழைதிருத்தி உருவாக்கத்தில்

உருபன் பகுப்பியின் பங்கு

ம. பார்கவி B.E., M.A, கி. உமா தேவி M.A

மொழியியல் ஆய்வுப்பிரிவுதமிழ்மொழித்துறை,

சென்னைப்பல்கலைக்கழகம்.

கணினியில் தட்டச்சு செய்யப்படும் தமிழ் உரைகளில் எழுத்துப்பிழைகள் ஏதேனும் காணப்பட்டால், அவற்றைக் கண்டறிந்து திருத்தும் தானியங்கு சொற்பிழைதிருத்தி உருவாக்கத்தில் உருபன் பகுப்பாய்வின் முக்கியத்துவம் பற்றிய கட்டுரையாக இது அமைகிறது.

சொற்பிழைகளுக்கான காரணங்கள்

தமிழில் தட்டச்சு செய்யும்போது, இரண்டு காரணங்களால் சொற்பிழைகள் ஏற்படலாம். ஒன்று தட்டச்சின்போது குறிப்பிட்ட எழுத்துக்கு உரிய விசைப்பொத்தானுக்குப் பதில் தவறாக மற்றொரு விசைப்பொத்தானை அழுத்திவிடுவதால் ஏற்படும் பிழைகள் (typographical errors). மற்றொன்று, எழுத்துபற்றிய குழப்பத்தினால் ஏற்படும் பிழைகள் (cognitive errors). எடுத்துக்காட்டாக, ஆகாயத்தைக் குறிக்கும் வானம் என்ற சொல்லைத் தட்டச்சு செய்யும்போது, “ன” எழுத்தா “ண” எழுத்தா என்று குழப்பம் ஏற்படலாம். இதன் விளைவாகவும் சொற்பிழைகள் ஏற்படலாம். “ந-ன-ண”, “ர-ற”, “ல-ள-ழ” ஆகிய எழுத்துகளில் பொதுவாகத் தட்டச்சு செய்யும்போது குழப்பம் ஏற்பட்டு, சொற்பிழைகள் நேரிடலாம்.

சொற்பிழைகளின் வகைகள்

சொற்பிழைகளில் இரண்டு வகைகள் காணப்படும். ஒன்று, ஒரு குறிப்பிட்ட பொருளைத் தரும் சொல்லுக்குப் பதிலாக, மற்றொரு சொல்லைத் தட்டச்சு செய்வதால் ஏற்படும் பிழையாகும். எடுத்துக்காட்டாக, “மல்லிகையில் நல்ல மணம் இருக்கிறது” என்ற தொடரில் “மணம்” என்ற சொல்லுக்குப் பதிலாக “மனம்” என்ற சொல்லைத் தட்டச்சு செய்வதால் ஏற்படும் பிழையாகும். இரண்டு சொற்களுமே தமிழில் உண்டு. இதுபோன்ற பிழைகளை ஆங்கிலத்தில் “Real Word Error” என்று அழைப்பார்கள். தமிழில் மெய்ம்மைச் சொற்பிழை என்று அழைக்கலாம்.

மற்றொரு வகை சொற்பிழைக்கு எடுத்துக்காட்டு, “உடல்” என்ற சரியான சொல்லுக்குப் பதிலாக “உடள்” என்று தவறாகத் தட்டச்சு செய்வதால் ஏற்படுகிற சொற்பிழையாகும். உடள் என்ற சொல்லே தமிழில் கிடையாது. இதுபோன்ற சொற்பிழைகளை ஆங்கிலத்தில் “No Word Error” என்று அழைப்பார்கள். தமிழில் இன்மைச் சொற்பிழை என்று அழைக்கலாம்.

சொற்பிழை திருத்தி

கணினியில் தட்டச்சு செய்யப்படும் உரையில் மேற்கூறிய மெய்ம்மை மற்றும் இன்மை சொற்பிழைகள் காணப்பட்டால், அவற்றைத் தானாகவே திருத்தித்தரும் பணியை மேற்கொள்ளும் மொழிக்கருவி (Auto Text Checking Tool) உருவாக்கப்படுவது மிகவும் தேவையான ஒன்றாகும். சொல்லாளர் மென்பொருள்களில் (Word Processors) இக்கருவிக்கு ஒரு முக்கியமான இடம் உண்டு. மைக்ரோசாப்ட் வேர்ட் மென்பொருளில் ஆங்கிலம் போன்ற மொழிகளுக்கு இந்தவகை மொழிக்கருவி உருவாக்கப்பட்டுள்ளது. தமிழுக்கு இல்லை.

தமிழுக்கென்று உருவாக்கப்பட்டுள்ள சில சொல்லாளர் மென்பொருள்களில் தமிழ்ச்சொற்பிழை திருத்திக் கருவிகள் உருவாக்கப்பட்டுள்ளன.

தமிழ்ச்சொற்பிழை திருத்தி உருவாக்கத்தில் - பிரச்சனைகள்

தமிழ் ஒரு ஒட்டுநிலைமொழியாகும் (Agglutinative language). தமிழில் உள்ள பெயர், வினை, வினையடை, பெயரடை போன்ற வேர்ச்சொற்கள் தொடர்களில் பயின்றுவரும்போது, தங்களது அகராதிப்பொருண்மையுடன் (Lexical meaning) சில இலக்கணப் பொருண்மைகளையும் (Grammatical meaning) ஏற்று வருகிறது. பெயர்ச்சொற்கள் எண், வேற்றுமை போன்ற இலக்கணப் பொருண்மைகளை ஏற்று வருகின்றன. வினைச்சொற்கள் காலம், எண், பால் போன்ற இலக்கணப் பொருண்மைகளை ஏற்றுவருகின்றன.

இலக்கணப் பொருண்மைகள் இயற்கைமொழிகளில் (Natural languages) இரண்டு வகைகளில் வெளிப்படுகின்றன. ஒன்று, ஒட்டுகள் அல்லது உருபுகள் (Affixes) மூலம் அவை வெளிப்படலாம். எடுத்துக்காட்டாக, தமிழ்ப்பெயர்ச்சொற்கள் வேற்றுமை ஏற்று வரும்போது, வேற்றுமைகளை வெளிப்படுத்தும் உருபுகளைத் தம்முடன் இணைத்துக்கொள்கின்றன. இவ்வாறு இலக்கண உருபுகளைத் தமிழ் வேர்ச்சொற்கள் (Root Words) ஏற்று, வேர்ச்சொல் + இலக்கண உருபுகள் என்று ஒட்டுநிலைச்சொற்களாகத் தொடர்களில் அமைகின்றன. ஆங்கிலம் போன்ற மொழிகளில் இலக்கண உருபுகள் வேர்ச்சொற்களோடு இணைந்தும் வருகின்றன, தனித்தும் வருகின்றன. ஆங்கிலத்தில் இறந்தகால விசுவிகள் வினைச்சொற்களோடு இணைந்து வருகின்றன. எ-கா. "Look - Looked".

வேற்றுமை உறவுகளை வெளிப்படுத்தும் இலக்கணச் சொற்களான முன் உருபு (Preposition), துணைவினைகளை (Auxiliary Verbs) வெளிப்படுத்தும் இலக்கணச்சொற்களான வினைநோக்கு (Modal), வினைப்பாங்கு (Aspectual) உருபுகள் தனித்து வருகின்றன. இதுபோன்று வேர்ச்சொற்களும் இலக்கணச் சொற்களும் தனித்தனியே பயின்றுவந்தால், அவற்றை நேரடியாக அகராதிகளின் துணைகொண்டு பிழை இருக்கிறதா இல்லையா என்பதைக் கண்டறியமுடியும். பிழைகளைத் திருத்தமுடியும்.

ஆனால், தமிழில் பொதுவாக அனைத்து இலக்கணச்சொற்களும் தனித்துவராமல், வேர்ச்சொற்களோடு இணைந்துதான் வருகின்றன. அகராதிகளில் இடம்பெறும் சொற்கள் வேர்ச்சொற்கள் தொடர்களில் இலக்கண உருபுகளோடு இணைந்துவருவதால், அவற்றில் எழுத்துப்பிழைகள் ஏதேனும் இருந்தால், அகராதியை மட்டும் வைத்துக்கொண்டு, பிழைகளைத் திருத்தமுடியாது. இலக்கண உருபுகளை அகராதிச்சொற்களிலிருந்து பிரித்து, பின்னர் அவற்றை வேர்ச்சொல் அகராதியிலும், இலக்கண உருபு அகராதியிலும் ஒப்பிட்டு, பிழைகளைக் கண்டறிந்து அவற்றைக் களையமுடியும்.

எனவே, தமிழ்மொழியைப் பொறுத்தமட்டில், தொடர்களில் பயின்றுவருகின்ற ஒட்டுநிலைச் சொற்களைப் பகுதி, விசுவிக் என்று பகுத்துப் பார்க்கும் தேவை உள்ளது. அதாவது, ஒட்டுநிலைச் சொற்களை உருபன் பகுப்பாய்வுக்கு (Morphological Analysis) உட்படுத்த வேண்டியுள்ளது.

இயற்கைமொழி ஆய்வு (Natural Language Processing - NLP)

மனித மூளையின் மொழிப்புலமானது இயற்கைமொழி அறிவைப் பெற்றுக்கொண்டு (அல்லது கற்றுக்கொண்டு), பொருள்பொதிந்த தொடர்களைப் புரிந்துகொள்ளவும் உருவாக்கவும் (Understanding and Generating) திறமையுடையதாக இருக்கிறது.

மனித மூளைபோன்று கணினியையும் இயற்கைமொழி அறிவைப் பெறவைத்து, மொழித்தொடர்களைப் புரிந்துகொள்ளவும் உருவாக்கவும் செய்யவைக்கும் முயற்சியே இயற்கைமொழி ஆய்வு என அழைக்கப்படுகிறது.

கணினியில் இயற்கைமொழி ஆய்வை மேற்கொள்ள மேற்கொள்ளப்படுகிற அல்லது உருவாக்கப்படுகிற வழிமுறைகள் (methods and formalisms) பற்றிய துறையே கணினிமொழியியல் (Computational Linguistics) என அழைக்கப்படுகிறது.

கணினிமொழியியல்

மனிதமூளையின் மொழிப்புலனில் இடம் பெற்றுள்ள குறிப்பிட்ட மொழியின் மொழி அறிவை எவ்வாறு கணினிக்கு அளிப்பது என்பது (Knowledge Representation) பற்றிய ஆய்வே கணினிமொழியியலின் அடிப்படையாக அமைகிறது. அதாவது மனிதமூளையின் மொழிப்புலனின் அறிவைக் கணினிக்கு எவ்வாறு அமைத்துக்கொடுப்பது என்பதே கணினிமொழியியலாகும். கணினிக்கு அந்த மொழி அறிவானது நிரல்வழிமுறைகளாகவும் (Algorithms) தரவு அமைப்புகளாகவும் (Data Structure) அளிக்கப்படுகிறது.

கணினிமொழியியலின் வளர்ச்சியின் பயனாகச் சொல்லாளர், பேச்சு - எழுத்துமாற்றி (Automatic Speech Recognizer - ASR), எழுத்து-பேச்சுமாற்றி (Text to Speech - TTS), தேடல்பொறி (Search Engine), இயந்திரமொழிபெயர்ப்பு (Machine Translation) போன்ற பல பயன்பாட்டு மொழிசார் மென்பொருள்கள் (Application Software) உருவாக்கப்படுகின்றன. இந்த உருவாக்கம் மொழித்தொழில்நுட்பம் (Language Technology) என்ற துறையாக வளர்ந்துள்ளது.

உருபன் பகுப்பாய்வு

தமிழ்த்தொடர்களில் பயின்றுவரும் ஒட்டுநிலைச் சொற்களைப் பகுத்து ஆய்வு மேற்கொள்ள கீழ்க்கண்டவை தேவையாகும்.

1. வேர்ச்சொல் தரவுத்தளம் (Root Word Database)
2. இலக்கண உருபு தரவுத்தளம் (Grammatical Affix Database)
3. வேர்ச்சொற்களும் இலக்கண உருபுகளும் இணையும் வரிசைமுறை (Morphotactics)
4. வேர்ச்சொற்களும் இலக்கண உருபுகளும் இணையும்போது ஏற்படும் உருபொலியன் மாற்றம் (Morphophonemics)

மனித மூளைக்குள் அமைந்துள்ள மொழிப்புலனில் (Language Faculty) மேற்குறிப்பிட்ட மொழி அறிவு (Knowledge of Language) புதைந்துள்ள காரணத்தால்தான், தமிழைத் தாய்மொழியாகக் கொண்ட எவரும் தமிழ்த்தொடர்களில் பயின்றுவரும் ஒட்டுநிலைச்சொற்களைப் பகுத்துப் பார்க்கும் திறன் படைத்துள்ளவர்களாக உள்ளனர் (competent to parse morphologically). அதாவது, அவர்களது மூளையில் உள்ள மொழிப்புலனில் உருபன் பகுப்பாய்வுக்குத் தேவையான மேற்குறிப்பிட்ட நான்குவகை மொழி அறிவுகளும் இடம்பெற்றுள்ளன.

உருபன் பகுப்பாய்வு இயற்கைமொழி ஆய்வுக்குப் பல வகைகளில் உதவி புரிகிறது. உயர்நிலை ஆய்வுகளான தொடரியல் ஆய்வு (Syntactic analysis), பொருண்மையியல் (Semantic analysis) ஆய்வு ஆகியவற்றிற்கு அடிப்படையாக உருபன் பகுப்பாய்வு அமைகிறது. ஒரு சொல்லை உருபன்களாக (அது தனிச்சொல்லாகவோ அல்லது ஒட்டுநிலைச்சொல்லாகவோ இருக்கலாம்) பகுத்து ஆய்வுசெய்யும் போதுதான் பெயர், வினை, பெயரடை, வினையடை போன்ற முதன்மை இலக்கண வகைப்பாடு (Grammatical Category) மற்றும் வினைமுற்று, வினையெச்சம், பெயரெச்சம் போன்ற உள்வகைப்பாடு (Sub-Category) பற்றிய விவரங்கள் தெரியவரும். இந்த விவரங்களுடன் சொற்கள், தொடரியியல் ஆய்வுக்கும் பொருண்மையியல் ஆய்வுக்கும் உள்ளீடாக (input) அனுப்பிவைக்கப்படவேண்டும். அதுபோன்று சந்தி விதிகளை முறையாகச் செயல்படுத்தவும் இந்த அறிவு தேவைப்படுகிறது.

தானியங்கு சொற்பிழைதிருத்தி (Auto spell-checker)

தானியங்கு சொற்பிழைதிருத்தியானது மொழித்தொழில்நுட்பத்தின் ஒரு படைப்பாகும். அதை உருவாக்கத் தேவையான அறிவானது, கணினிமொழியியல் துறையில் உருவாக்கப்படுகிறது. தமிழ்ச்சொற்பிழை திருத்தி உருவாக்கத்தில் உருபன் பகுப்பியின் (Morphological Analyser) பங்கு மிக அடிப்படையானது. தொடர்களில் பயின்றுவரும் ஒரு சொல் வேர்ச்சொல்லாக இருந்து, அவற்றில் எழுத்துப்பிழை இருந்தால், ஒரு நல்ல மின்அகராதியின் (Electronic Dictionary) துணையுடன் ஒப்பிட்டு, சரியான சொற்களைப் பதிலீடு (Replacement) செய்யமுடியும். அச்சொல் வேர்ச்சொல்லாக இல்லாமல், ஒட்டுநிலைச்சொல்லாக இருக்குமேயானால், அகராதியின் துணைமட்டும் போதாது. முதலில் அச்சொல்லை உருபன் பகுப்பாய்வுக்கு உட்படுத்தி, அகராதி வேர்ச்சொல், இலக்கண உருபுகள் என்று பகுத்துக் கண்டறிவது தேவையாகிறது. முன்பே கூறியவாறு இதற்கு நான்குவகை மொழியறிவு தேவைப்படுகின்றன. இந்த நான்குவகை மொழியறிவையும் கணினிக்கு அளித்து, ஒட்டுநிலைச் சொல்லைப் பகுத்து ஆய்வு செய்யவேண்டும். இதுபோன்ற ஒரு உருபன்பகுப்பாய்வானது தொடரில் ஒரு சொல் பிழையாக அமைந்துள்ளதா இல்லையா என்பதைச் சோதித்து அறிய உதவுகிறது.

“படித்தான்” என்ற வினைமுற்றுச்சொல் ஒன்றில் எழுத்துப்பிழை இருக்கிறதா இல்லையா என்பதை ஆய்ந்தறிய வேண்டும் என்று வைத்துக்கொள்வோம். அச்சொல் முதலில் அகராதியில் இடம் பெற்றுள்ள விகுதி ஏற்காத தனிச்சொல்லா என்று அகராதியுடன் ஒப்பிட்டுப்பார்க்கவேண்டும். அகராதியில் இடம் பெற்றிருந்தால், அதை உருபன் பகுப்பிக்கு அனுப்பிவைக்கத் தேவையில்லை.

அகராதியில் இடம் பெறவில்லை என்றால், இரண்டு வகைகளில் சிந்திக்கலாம். ஒன்று, தனிச்சொல்லாக இருந்தும், எழுத்துப்பிழை உடைய சொல்லாக இருக்கலாம். எடுத்துக்காட்டாக, “தறை” (“தரை” என்பதற்குப் பதிலாக) என்ற ஒரு தவறான சொல்லை அகராதியுடன் ஒப்பிடும்போது, அது அங்கு இடம் பெற்றிருக்காது. எனவே அது எழுத்துப்பிழை உடைய தனிச்சொல்லாகவும் இருக்கலாம். அச்சொல்லில் சில எழுத்துகளை மாற்றியமைத்து, அகராதியுடன் ஒப்பிட்டுப்பார்க்கலாம். அப்போது அச்சொல் இடம் பெற்றிருந்தால், அதைப் பரிந்துரையாகக் காட்டலாம்.

“படித்தான்” என்ற ஒரு ஒட்டுநிலைச் சொல்லை அகராதியுடன் ஒப்பிடும்போதும் அது இடம் பெற்றிருக்காததால், மேற்குறிப்பிட்ட வகையில் எழுத்துகளை மாற்றியமைத்து அகராதியுடன் ஒப்பிட்டு பார்க்கலாம். அப்போதும் அது அகராதியில் காணப்படவில்லை என்றால், அது ஒரு ஒட்டுநிலைச் சொல்லாக இருக்கலாம் என்ற ஒரு முடிவுக்கு வரலாம். அச்சொல்லை உருபன் பகுப்பாய்விற்கு அனுப்பவேண்டும். உருபன் பகுப்பியானது அச்சொல்லை “படி + த்த் + ஆன்” என்று பகுத்துவிட்டது என்றால், எழுத்துப்பிழையில்லாத ஒரு ஒட்டுநிலைச்சொல் என்று எடுத்துக்கொள்ளலாம். மாறாக, பகுக்கமுடியவில்லை என்றால், அச்சொல்லில் ஏதேனும் பிழை இருக்கலாம் எனக் கருதலாம். எடுத்துக்காட்டாக, “படித்தான்” என்பதற்குப் பதிலாக படித்தான்* என்று தட்டச்சு செய்திருந்தால், உருபன் பகுப்பியால் இதைப் பகுக்கமுடியாது. எனவே எழுத்துப்பிழை உடைய சொல் இது என்று முடிவுக்கு வரலாம்.

பின்னர் சில எழுத்துகளை மாற்றியமைத்து, உருபன் பகுப்பிக்கு அனுப்பலாம். அப்போது அச்சொல் பகுக்கப்பட்டது என்றால், மாற்றியமைக்கப்பட்ட சொல்லே எழுத்துப்பிழை இல்லாத சொல் என்று முடிவுக்கு வரலாம். எடுத்துக்காட்டாக, மேற்குறிப்பிட்ட சொல்லில் (படித்தான்*) “-ண்” என்பதை “-ன்” என்ற எழுத்தால் மாற்றியமைக்கும்போது, உருபன் பகுப்பி இச்சொல்லைப் பகுக்கிறது.

ஒரு சொல்லை மேற்கூறியவாறு மாற்றியமைத்தபிறகும் உருபன் பகுப்பியால் பகுக்க இயலவில்லை என்றால், அச்சொல் தற்போது மாற்ற இயலாத எழுத்துப்பிழை உடைய சொல்லாக இருக்கலாம். அல்லது அகராதியில் இடம்பெறாத ஒரு சொல்லாக இருக்கலாம். ஒரு சொல் எழுத்துப்பிழை இல்லாத சொல்லா இல்லையா (தனிச்சொல் அல்லது ஒட்டுநிலைச்சொல்) என்பதைக் கண்டறிய இவ்வாறு உருபன் பகுப்பி பயன்படுகிறது. எழுத்துப்பிழைகளைத் திருத்தவும் பயன்படுகிறது.

தானியங்கு சந்திப்பிழைதிருத்தி உருவாக்கத்தில்

உருபன் பகுப்பியின் பங்கு

R. பத்ம மாலா, MCA .,(Ph.D.), M. பார்வதி B.E. M.A

மொழியியல் ஆய்வுப்பிரிவு , தமிழ்மொழித்துறை,
சென்னைப் பல்கலைக்கழகம்

கணினியில் தமிழைத் தட்டச்சு செய்யும்போது தானாக சந்திப்பிழைகளைச் சரிசெய்யும் கணினிநிரல் (Computer Program) தானியங்கு சந்திப்பிழைதிருத்தியாகும்(Auto Sandhi Checker). இதன் உருவாக்கத்தில் உருபன் பகுப்பியின் (Morphological Analyser) பங்கை ஆராய்வதே இக்கட்டுரையின் நோக்கம்.

உருபொலியனியல் மாற்றங்கள் (Morphophonemic Changes)

தமிழில் நிலைமொழி இறுதியெழுத்தும் வருமொழி முதலெழுத்தும் சேரும் போது ஏற்படும் மாற்றங்களே உருபொலியனியல் மாற்றங்கள் ஆகும். இதனையே சந்தி என்கிறோம். ஒரு வேர்ச்சொல்லோடு விசுதிகள் சேரும் சொற்சந்தியும், சொற்களோடு சொற்கள் சேரும் தொகைச்சந்தியும் அகச்சந்தி எனப்படும். சொல்லுக்குள் இல்லாமல் தொடரில் இரண்டு சொற்களுக்கு இடையில் ஏற்படும் மாற்றங்கள் புறச்சந்தி எனப்படும்.

சந்திப்பிழை சாத்தியக்கூறுகள் :

ஒரு தொடரில் பின்வரும் நிலைகளில் சந்திப்பிழைகள் ஏற்பட வாய்ப்புகள் உண்டு.

- ஒற்று மிக வேண்டிய இடத்தில் இடாதிருத்தல் எ.டு. படித்து பார் *.
- ஒற்று மிகக்கூடாத இடத்தில் இடல் எ.டு. அழகானப் பெண் *
- ஒற்று மிக வேண்டிய இடத்தில் தவறான ஒற்றை இடல் எ.டு. நடித்துக் பார்*.

உருபொலியனியல் மாற்றத்தின் முக்கியத்துவம்

பொருள் மயக்கத்தை(Lexical Ambiguity) நீக்கிப் பொருளைத் தெளிவாகப் புரிந்துகொள்வதற்கு உருபொலியனியல் மாற்றங்கள் துணைபுரிகின்றன. எடுத்துக்காட்டாக,

- தந்த பலகை; தந்தப் பலகை
- தயிர் கடை; தயிர்க் கடை

என்பனவற்றில் 'தந்த பலகை' என்பது ஒருவர், இன்னொருவருக்குத் தந்த பலகை என்றும் 'தந்தப் பலகை' என்பது தந்தத்தினாலான பலகை என்றும் பொருள் வேறுபாட்டைப் புரிந்துகொள்ள, சந்தி தேவைப்படுகிறது. இதேபோன்று, 'தயிர் கடை' என்பது தயிரைக்கடை என்ற பொருளையும், 'தயிர்க் கடை' என்பது தயிரையுடைய கடை என்ற பொருளையும் கொண்டுள்ளது.

சந்திக்கான பல்நிலை அறிவு (Multilevel knowledge):

தமிழ்மொழியின் சந்தி மாற்றங்களைத் தெளிவாகத் தெரிந்துகொள்ள ஒலியனியல், உருபனியல், தொடரியல் மற்றும் பொருண்மையியல் அறிவு போன்றன இன்றியமையாதவை.

தெரு+இல் = தெருவில்

என்பதில் நிலைமொழி இறுதியில் முன்னுயிர் (Front Vowel) உகரமும் வருமொழி முதலில் உயிரும் இருப்பதால் வகர உடம்படுமெய்(Glide) தோன்றியுள்ளது. இது ஒலியனியல் (Phonology) அடிப்படையிலான மாற்றம்.

அழகான பையன் தெளிவாகப் பேசினான்

இத்தொடரில் அழகான, தெளிவாக என்ற நிலைமொழிகளின் இலக்கண வகைப்பாட்டைக்கொண்டுதான் (Grammatical Category) சந்தி மாற்றம் செய்ய முடியும். இது உருபனியல் (Morphology) அடிப்படையிலான மாற்றம்.

- மாடு+கன்று=மாடுகன்று
- மாடு+கன்று=மாட்டுக்கன்று

மேற்கூறிய எடுத்துக்காட்டுகளில் நிலை, வருமொழிகள் ஒன்றையெனினும் இரண்டு விதமான சந்தி மாற்றங்கள் ஏற்படுகின்றன. மாடு மற்றும் கன்று எனும்போது மாடு கன்று என்றும், மாட்டினுடைய கன்று என்று வேற்றுமை உறவுடையதெனில் மாட்டுக்கன்று என்றும் மாற்றமடைகின்றன. இது தொடரியல் அடிப்படையிலான (Syntax) சந்தி மாற்றம்.

- பழம்+கூடை=பழக்கூடை
- பழம்+கூடை=பழங்கூடை

பழங்களையுடைய கூடையெனில் பழக்கூடை என்றும், பழமையான கூடையெனில் பழங்கூடை என்றும் மாற்றமடைகின்றன. இது பொருண்மையியல் (Semantics) அடிப்படையிலான சந்தி மாற்றம்.

உருபன் பகுப்பாய்வு

சொற்களின் இலக்கண வகைப்பாட்டை நாம் பல்வேறு நுண்மையான இலக்கண அறிவு (Grammatical knowledge) மற்றும் உலகியல் அறிவின் (Pragmatic knowledge) துணையோடு அறிகிறோம். மனித மூளைக்கு இது சாதாரணமே! ஆனால், இவ்விலக்கண வகைப்பாட்டைக் கணினிக்கு எவ்வாறு கற்பிப்பது? தானாக ஒரு சொல்லின் இலக்கண வகைப்பாட்டை அறிவதற்குக் கணினிக்கு எவ்வகையான அறிவைத் தரவேண்டும்? தமிழில் இடம்பெறும் தொடர்களில் வேர்ச்சொற்கள் தனித்தும், விகுதிகளேற்றும் காணப்படுகின்றன. தனித்த வேர்ச்சொற்களைக் கண்டறிவதற்கு அகராதினைப் பயன்படுத்தலாம். ஆனால், விகுதியேற்ற சொற்கள் அகராதிகளில் இடம்பெறாது.உள்எீட்டுச் சொற்கள் (Input words) அனைத்தையும் பகுப்பாய்விற்கு உட்படுத்தித்தான் அவற்றின் இலக்கண வகைப்பாட்டைக் கணினியால் பெறமுடியும். எனவே, உருபன் பகுப்பாய்வி அவசியமாகிறது.

தானியங்கு சந்திப்பிழைதிருத்தி (Auto Sandhi Checker)

மனித மூளையில் பொதிந்துள்ள மொழியறிவை மின்னணு வடிவில் கணினிக்கு வழங்கி, மனிதன் இயற்கை மொழிகளைப் புரிந்துகொண்டு கையாளுவதைப்போன்று கணினியையும் செய்யவைப்பதே இயற்கை மொழியாய்வின் (Natural Language Processing) நோக்கம். இதற்கான தரவு அமைப்பும் (Data Structure), நிரல்வழிமுறை (Algorithm) உருவாக்கமும் இணைந்ததே கணினி மொழியியல் (Computational Linguistics). இக்கணினிமொழியியலின் துணையோடு மொழிக்குத் தேவையான பல பயன்பாட்டு

மென்பொருள்களை உருவாக்க உதவும் நுட்பமே மொழித்தொழில்நுட்பம் (Language Technology). தானியங்கு சந்திப்பிழைதிருத்தி என்பது மொழித்தொழில்நுட்பத்தினால் உருவாக்கப்பட்ட ஒரு மொழிக்கருவி (Language Tool). தமிழுக்கான தானியங்கு சந்திப்பிழைதிருத்தி உருவாக்கத்தில் சந்தி விதிகளை முறையாகச் செயல்படுத்துவதற்குப் பின்வரும் இலக்கண அறிவு தேவைப்படுகின்றன.

- வருமொழியின் முதலெழுத்து
- நிலைமொழியின் இறுதியெழுத்து
- நிலைமொழியின் அசையமைப்பு (Syllabic Structure)
- நிலைமொழியின் இலக்கண வகைப்பாடு (Grammatical Category)
- உருபொலியனியல் மாற்றங்கள் (Morphophonemic Changes)
- உருபங்களின் வரிசைமுறை (Morphotactics)

தொடர்ச்சந்தியில் சந்திமாற்றங்களைச் செயல்படுத்த முதலில் வருமொழி முதலெழுத்து சந்தி வல்லினமாக (க்,ச்,த்,ப்) இருக்க வேண்டும். நிலைமொழி இறுதியெழுத்து சந்தியிடத் தேவையான எழுத்துகளாக இருக்க வேண்டும்.

எ.கா. அவனே படித்தான்

என்ற தொடரில் நிலைமொழியின் இறுதியெழுத்துக்குச் சந்தி தேவையில்லை. இதற்கு மேற்கூறிய முதல் இரு இலக்கண அறிவு தேவைப்படுகின்றன. நிலைமொழியின் இறுதியெழுத்து சந்திமாற்றத்திற்குரியதாகவும், வருமொழி முதல் சந்தி வல்லினமாகவும் இருந்தால், நிலைமொழி வேர்ச்சொல்லா என்பதை அகராதியுடன் ஒப்பிட்டுப் பார்த்து அறியவேண்டும். அகராதிச்சொல் இல்லையென்றால் அதை ஒட்டுநிலைச் சொல்லாகக் கருதி அதை உருபங்களின் வரிசைமுறையின் அடிப்படையில் உருபன் பகுப்பாய்வுக்கு உட்படுத்த வேண்டும்.

உருபன் பகுப்பாய்வுக்கு உள்ளீடாக வழங்கப்பட்ட ஒரு சொல் வேர்ச்சொல் + ஒட்டுகளாகப் பகுக்கப்படுகின்றன. இங்கு வேர்ச்சொல்லைக் கண்டறிவது முக்கியமான ஒன்றாகும். உருபன் பகுப்பாய்வுமூலம் வேர்ச்சொல்லைப் பெற்றபிறகுதான், அதன் அசையமைப்பைக் கண்டுபிடித்து வேர்ச்சொல் மற்றும் ஒட்டுகளுக்கிடையே சந்தி மாற்றங்களைச் செயல்படுத்தலாம்.

நிலைமொழி ஆகார ஈற்று ஓரசைச்சொல்லாகவும், உகரயீற்று நிரையசைச்சொல்லாகவும் இருந்து, வருமொழி பன்மை விகுதி 'கள்' ஆகவும் இருந்தால் ஒற்றுமிகும்.

எ.டு. புறா + கள் = புறாக்கள்

நிலைமொழியின் அசையமைப்பு மெய்-உயிர்(குறில்)-மெய் (CVC) அல்லது உயிர்(குறில்)-மெய் (VC) ஆக இருந்தால் இறுதிமெய் இரட்டிக்கும்.

எ.டு. கல் + ஐ = கல்லை

மேலும் உருபன் பகுப்பாய்வு மூலம் பெற்ற வேர்ச்சொல்லின் முதன்மை வகைப்பாட்டை (Grammatical Category) அகராதி உதவியுடன் பெறலாம். வேர்ச்சொல் மற்றும் பகுக்கப்பட்ட விகுதிகளைக் கொண்டு உள்ளீட்டுச் சொல்லின் உள்வகைப்பாட்டை (Sub-Category) அறியலாம்.

எ.கா. படித்த பையன் / படிக்கப் பார்த்தான்

இங்கு, படித்த, படிக்க என்ற இரு நிலைமொழிகளும் படி என்ற வினைச்சொல்லிருந்து தோன்றியவையே. இருப்பினும், படித்த என்பதன்பின் ஒற்று மிகவில்லை. படிக்க என்பதன்பின் ஒற்று மிகுகிறது. ஏனெனில், இவ்விரண்டு சொற்களின் முதன்மை வகைப்பாடு வினையாக இருந்தாலும் படித்த என்பதன் உள்வகைப்பாடு பெயரெச்சமாகவும் படிக்க என்பதன் உள்வகைப்பாடு செய வாய்பாட்டு

வினையெச்சமாகவும் உள்ளன. எனவே, தானியங்கு சந்திப்பிழைதிருத்திக்குச் சொற்களின் முதன்மை வகைபாடும்படும் போதாது. சொற்களின் உள்வகைப்பாடும் (Subcategory) பெரிதும் தேவைப்படுகின்றது. உள்வகைப்பாட்டைத் தெளிவாக அறிவதற்குச் சொற்களை நுண்மையான பகுப்பாய்விற்கு உட்படுத்தல் அவசியம்.

இவ்வாறு செம்மையான தானியங்கு சந்திப்பிழைதிருத்தி உருவாக்கத்திற்கு ஆதாரங்களாக விளங்கும் சொல்லின் அசையமைப்பைக் கண்டறிதல் , உருபங்களின் வரிசைமுறையைத் தீர்மானித்தல், சொற்களின் வகைப்பாடு மற்றும் உள்வகைப்பாட்டை அறிதல் ஆகியவற்றில் உருபன் பகுப்பியின் பங்கு இன்றியமையாதது.

Spell Checker for Tamil using Finite State Automata

Anitha.S Pillai

Hindustan University
anitha2293@gmail.com

Abstract

The problem of detecting and correcting misspelled words in text has received great attention due to its importance in several applications like text editing systems, optical character recognition systems, morphological analysis and tagging. Other applications like machine translation and information extraction, operate on text. There are possibilities that there may be errors in these text. The ability to automatically detect and correct spelling error should be of great help to those applications. The problem of detecting and correcting misspelled words in text is usually solved by checking whether a word already exists in the dictionary or not. If not, we try to extract words from the dictionary that are most similar to the word in question. This will not work for all languages where a large number of words can be derived from the root word and it is not feasible to store all the words in the dictionary. Hence a Morphological analyzer also plays an important role in these languages.

In this paper an approach for automatic correction of spelling mistakes in Tamil document using the Finite State Automata (FSA) is proposed.

Introduction to spell checking

This paper discusses the stages involved in the development of a Tamil Spell Checker. Since in Tamil, a large number of words can be derived from a root word, a purely dictionary based approach for Spell Checking is not practical. Hence a 'Rule cum Dictionary' based approach is followed. The lexicon (dictionary) is stored in the form of finite state automata.

The different modules in the Spell Checker Engine viz. Morphological Analyzer, Error detection and suggestion generation module are also explained.

A Spell checker is a tool that will check the spelling of words in a document, validate them and in case the spell checker finds error, list out the correct spelling in the form of suggestions. Any word processor should have a spellchecker associated with it as the user may commit mistakes during typing. According to Damerau (1964), 80% of all misspelled words in a sample of human keypunched text were caused by single-error misspellings. It could be either due to insertion, deletion, substitution or transposition. Kukich (1992) breaks down human typing errors into two classes. Typographic errors are generally related to keyboard. Cognitive errors are caused by writers who don't know how to spell the word.

The spellchecker detects the mistakes and prompts the user with a set of suggestions, which will aid the correction of the misspelled word. The approach followed in the design of an Indian language Spell Checker has to be different from those for Roman scripts, because in Roman alphabet each letter is

complete by itself and represents a sound where as in Indian languages the characters are more syllabic in nature and most of the consonants will have a vowel

sound added to it. Unlike other languages, development of Spellcheckers for Indian languages, especially Dravidian languages like Malayalam, Tamil, Kannada and Telugu is a bit complicated.

Here the suffixes, postpositions and case endings agglutinate with the verbs, nouns, adverbs or pronouns. Also one or more suffixes can combine with the base word. In other Indian languages, for eg. Hindi, case ending will not agglutinate with the base form. Also combinations like verb-verb, noun-noun, verb-noun etc. are not permitted in many other Indian languages.

Hence Morphological analysis of the input word is a must for Tamil Spell checker. Tamil is a morphologically rich language in which most of the morphemes coordinate with the root words in the form of suffixes. Person, gender and number markings combine with the root words. For error detection and suggestion an efficient morphological analyzer is required.

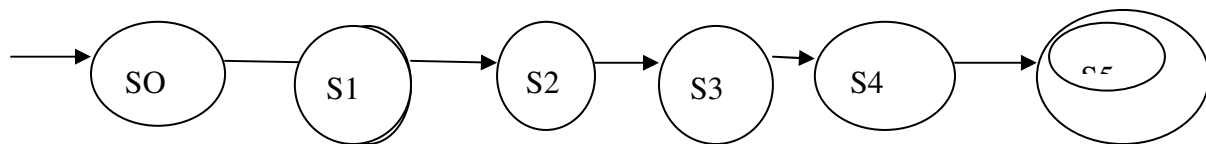
Spell checker application presents valid suggestions to the user based on each mistake they encounter in the user's document. The user either selects from the suggestions or choose to ignore the suggestions and accept the current word as correct.

Finite State Automaton

Automaton is represented as a directed graph: a finite set of vertices (nodes), together with a set of directed links between pairs of vertices called arcs. Each node corresponds to a state. States are represented as circles with name tags in them. Arcs are represented by arrows going from one state to another state. The final states are represented by double circles.

The machine starts at the initial state, runs through a sequence of states by computing a morpheme in each transition, and ends in the final state. The path moves from the initial point on the left to the final point on the right, proceeding in the direction of arrows. Once the arrow moves one step, there is no backward movement (Of course, recursion of an item can be shown by using closed loops). The resulting Finite State Automata (FSA) is deterministic in the sense that given an input symbol and a current state, a unique next state is determined. A deterministic finite state automaton (DFA) is perhaps the simplest type of machine that is still interesting to study.

However, it is an enormously useful practical abstraction because DFAs still retain sufficient flexibility to perform interesting tasks, yet the hardware requirements for building them are relatively minimal. DFAs are widely used in text editors for pattern matching, in compilers for lexical analysis, in web browsers for html parsing, and in operating systems for graphical user interfaces. They also serve as the control unit in many physical systems including: vending machines, elevators, automatic traffic signals, and computer microprocessors. They also play a key role in natural language processing and machine learning.



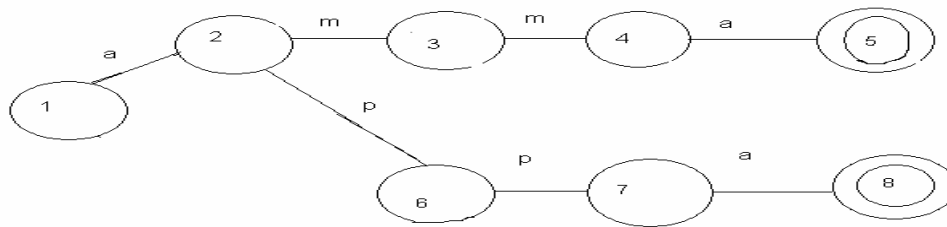
Structure of Finite State automaton

Here S0 represents initial state, S1,S2,S3 and S4 represents intermediate state and S5 represents final state. The automaton has five states which are represented by nodes in the graph. State S0 is the start state which we represent by the incoming arrow. State S5 is the final state or accepting state which we represent by the double circle.

Representation of words using Finite State Automaton

Tamil lexicon is stored as a Finite state Automaton. There is an initial state and final state for each word. The set of valid characters that help to move from the initial state to final state is specified for all the Tamil root words.

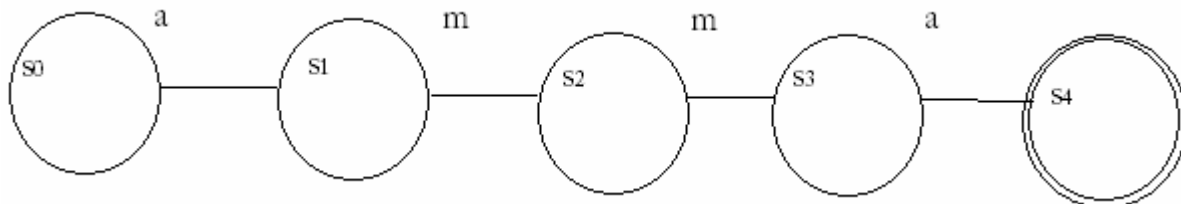
Eg: initially the system is in state 1(initial state) .when a character 'a' is input it goes to state2. From state2 on receiving the input 'm' it goes to state 3. The list of characters that characters that help 'a ' to move from the initial state to final state is given.



Representation of words amma and appa using FSA

Checking of Words

It starts at the initial state (S0), checks the next morpheme of the input. If it matches the symbol on an arc leaving the current state, then it crosses that arc, and moves to the next state, and thus, advances one symbol in the input. Such a process gets iterated until the machine reaches the final state, successfully recognizing all the morphemes in the input string. But if the machine gets some input that does not match an arc, then it gets stuck there and never gets to the final state. This is considered as the FSA / machine rejecting or failing to accept an input.



The word 'amma' is accepted by the finite state automata

Word not recognized by Finite state automaton

If a set of input characters do not help to move from the initial state to final state then the automata will not be recognized for the given word. This could happen due to 2 reasons:

- 1) The word is an inflected word i.e. suffixes are attached to the root word.
- 2) Spelling mistake in the word

Inflected Word : Root words to which suffixes are attached. These words will not be available in the lexicon. So Morphological analyzer is required to segment the word to root and suffix.

Morphological analysis: There will be certain suffixes along with the root word. This word can be called as an inflected word. The respective inflected word is stripped to its corresponding root word and suffix using backtracking algorithm such that the whole word can be written as root word+suffix. But in some cases when the suffix is stripped off the input string the root word will not be present in the root database. In such cases, we use orthographic rules which are solely used for certain words which are meant to be modified based on these rules.

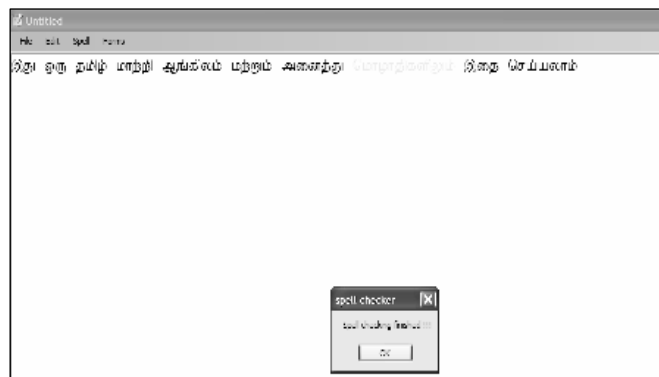
For example:-pasangal. In this word the 'gal' denotes the plural form. After the plural suffix is removed the rest of the word is checked with the root database. If the word is not present in the root database then orthographic rules are applied such that the input string is identified as correct.

Sandhi checker

The Sandhi checker deals with the orthographic changes that occur in a root word when suffixes are added to it . When the input string has suffixes added to it and they are stripped off their respective suffixes, in some cases the root word may not be present in the root database. For example:- consider the word 'pazhangal'. Once the suffix 'gal' is stripped off the rest of the word shows 'pazhang' which is not present in the root database. Here the sandhi checker helps in finding out whether the word is right or not. And finally the word is displayed as a correct word.

Suggestions

The minimum edit distance (or just edit distance) between two strings is the least number of elementary editing operations - insertions, deletions, and substitutions - that are needed in order to transform the first string into the second one.



Screen Layout of Spell Checker

For example, the edit distance between the strings summer and summary is 2: we substitute the e for a, and then we insert y at the end. No other shorter edit sequence exists. Therefore, the edit distance is 2. These metric measures how similar two strings are. Edit distance is also symmetric: if ed is a binary function mapping two strings to their edit distance, then, for every two strings s and t, ed(s, t) is always equal to ed(t, s).

The whole program is developed using visual basic.net 2008 and MS-Access database. When the user types a word and if the word cannot be represented as a FSA an error is shown. This means that this particular word is not a correct one and so the FSA fails to move from the initial state to the final state. In such cases using minimum edit distance possible suggestions are displayed to the user to choose from. When the user types the word and press the space bar to type the next word, if there is an error then the word typed changes to blue. When the user right-click the mouse, suggestions are provided to correct the misspelled word. The suggestions are based on minimum edit distance.

Conclusions:

Users can type text in Tamil and whenever the application encounters a misspelled word it highlights the error by changing the font colour of the wrong word to yellow. When we right click on this wrong word a list of possible suggestions are displayed.

References:

1. Thomas Lehmann (1993) "A Grammar of modern Tamil", Second Edition, Pondicherry Institute of Linguistics and Culture.
2. Dr. K. Balasubramanian (2001), "Studies in Tholkappiyam", Annamalai University.
3. A Smart Spell checker system <http://www.coe.neu.edu/>
4. Anandan. P, Ranjani Parthasarathy, Geetha T.V. (2001), Morphological Analyzer for Tamil, ICON 2002, RCILTS-Tamil, Anna University, India.
5. D.Jurafsky and J.H.Martin "Speech and Language Processing", Prentice Hall 2000
6. Yo-Sub Hana, Derick Woodb Obtaining shorter regular expressions from finite-state automata Theoretical Computer Science 2006

தமிழ் மொழிக்கருவிகளைக் கணிப்பொறியாக்கம் செய்வதின் தேவைகளும் தரப்படுத்துதலின் இன்றியமையாமைகளும்

Language Tools for Tamil : Needs for Computerization and Necessities for Standardization

ச.பாஸ்கரன்

துறைத்தலைவர்

கணிப்பொறி அறிவியல் துறை

தமிழ்ப் பல்கலைக்கழகம்

தஞ்சாவூர் -614 010

sbaskarantj@yahoo.com

1.0 முன்னுரை

ஆல்பர்ட் ஐன்ஸ்டீன் [Albert Einstein] காலத்தையும் நேரத்தையும் ஒரே மாறியாக [Variable] வரையறை செய்தது இயற்பியல் மற்றும் தொழில் நுட்பத்தில் பெரும் புரட்சியை ஏற்படுத்தியது. இதன் பயனைச் சமுதாயம் துய்த்தது. தற்போது இணையம் [Internet] காலத்தையும் நேரத்தையும் மறுவரையறை [Redefinition] செய்துள்ளது. இந்த வரையறை இதுவரை இல்லாத புது நடையியல் மாற்றத்தைத் தகவல் பரிமாற்றம், பணித்தன்மை, கல்வி, ஆய்வு, வணிகம், பொழுதுபோக்கு, ஓய்வு, உடல் நலம் போன்றவற்றில் உருவாக்கியுள்ளது. உலகத்தைக் கைக்கொண்டுள்ள இணையம் பல தடைகளைத் தகர்த்தெறிந்துள்ளது. அறிவியல் கருவியான கணிப்பொறியை மக்கள் கருவியாக மாற்றிவிட்டது. இதற்கு வங்கிகளின் ஏ.டி.எம்-கள் [ATM] நல்ல சான்று. இத்தகைய மாற்றத்தை ஆல்வின் டோஃப்ளர் [Alvin Toffler] மூன்றாவது அலை என்றும், பீட்டர் டிரக்கர் [Peter Drucker] தொழிற் புரட்சியின் பிற்காலம் என்றும் அழைக்கத் தொடங்கினர். நாளடைவில் பொருளாதார, மேலாண்மை, எதிர்காலவியல் அறிஞர்கள் பலரால் 'தகவல் யுகம்' என்றழைக்கப்படுகிறது. இந்த யுகத்தில் 'தகவல்'; அதைத் தாங்கி நிற்கும் அறிவும்தான் பெருஞ்செல்வம் என மதிக்கப்படுகிறது. இந்நிலைதான் உலகத்தை வழிநடத்திச் செல்கிறது. இப்புது யுகத்தின் மாற்றத்திற்கு உட்படுத்திக்கொள்ளாத சமுதாயமும் மொழியும் வளர்ச்சிப் பாதையில் பயணிப்பதற்கான வாய்ப்புகளை இழக்க நேரிடும்.

'தகவல் யுகத்தில்' கணிப்பொறிதான் கருவி. தகவலுக்கு மொழிதான் கருவி. எனவே, மக்கள் பேசும் மொழி கணிப்பொறிக்குத் தேவை. உலக மொழிகளைக் கணிப்பொறியில் கையாள புத்தம் புதிய வசதிகளும் தொழில்நுட்பங்களும் வளர்ந்த வண்ணம் உள்ளன. தமிழ் விசைப்பலகையின் வடிவமைப்பைத் தமிழக அரசு தரப்படுத்தியது. இதுபோல், மொழி ஆய்வுக் கருவிகளான மின்-அகராதி [E-Dictionary], சொற்களஞ்சியம் [Thesaurus], பெருந்தரவு [Corpus] போன்றவைகளை உருவாக்க வேண்டுவதும், தரப்படுத்த வேண்டுவதும் அவசியமாகும். தரப்படுத்துவதற்குத் தவறும் அல்லது தாமதிக்கும் நிலையில், விசைப்பலகையின் வடிவமைப்பைத் தரப்படுத்த அல்லலப்பட்டதுபோல் தடுமாற்றம் நிகழ வாய்ப்புள்ளது. இவை பற்றிய செய்திகளை எடுத்துரைப்பதும் கருத்தளிப்பதும் இக்கட்டுரையின் நோக்கமாகும்.

2.0 கணிப்பொறியில் உலக மொழிகள்

அவரவர்கள் மொழியிலுள்ள தகவல்களை கணிப்பொறியில் கையாளுவதற்கான வசதிகளை உருவாக்கிக்கொள்ள வேண்டியது அந்தந்த மொழியாளர்களின் கடமை என்ற நிலை இருந்தது. இந்நிலை மாறி, உலகப் பயன்பாட்டிலுள்ள மென்பொருள் தயாரிப்பு நிறுவனங்களின் கடமை இது என்றாகியுள்ளது. இதன் பயனாய் ஒருங்குறி [Unicode] எழுத்துரு உருவாகியது. மைக்ரோ சாஃப்ட் நிறுவனத்தின் XP இயக்க ஒருங்கின் [OS]கீழியங்கும் அலுவலகத் தொகுப்பிலுள்ள [M.S.Office Package] மென்பொருள்கள் உலகத்திலுள்ள பல மொழிகளின் எழுத்துக்களை ஆதரிக்கின்றன. இருப்பினும் மொழிச்செய்திகளைக் கணிப்பொறியில் உள்ளீடு செய்வதற்கான முறைமைகளை வரையறை செய்வதும், தரப்படுத்துவதும் அந்தந்த மொழியாளர்களின் பொறுப்பாக உள்ளது. இந்த அடிப்படைத்தேவையிலிருந்து அண்மைக்காலத் தொழில் நுட்ப மேம்பாடுகளுக்கேற்ற வண்ணம் உருவாக்க வேண்டிய தமிழ் மொழிக்கருவிகளின் தேவை வரையிலான சில விவரங்களைச் சுருங்கக் காணலாம்.

3.0 மொழிக்கருவிகள்

தகவல் தொழில் நுட்பக்கருவியாக அமைந்துள்ளதால் கணிப்பொறியில் பல மொழிகளின் பயன்பாடு பன்மடங்கு வளர்ந்துள்ளது. எழுத்துப்பிழைத்திருத்தி [Spell Checker], இலக்கணப்பிழைத்திருத்தி [Grammar Checker], அகரவரிசையாக்கம் [Sorting] போன்ற பணிகளுக்கெல்லாம் மொழிக் கருவிகளே [Language Tools] அச்சாணியாக விளங்குகின்றன. இப்பயன்பாடுகளோடு உருவாக்கப்படும் மென்பொருளொன்றின் குறை/நிறைகள், வெற்றி/தோல்விகள், பயன்பாட்டுத்தரம், புகழ், பரவலாக்கம் போன்றன மென்பொருள் தன்னகத்தே உள்ளடக்கியுள்ள அல்லது பொதிந்துள்ள மொழிக்கருவிகளைப் பொறுத்தே அமைகின்றன. கணிப்பொறித் தமிழ் பயன்பாடுகளுக்கும் ஆய்வுகளுக்கும் தேவைபடும் மொழிக்கருவிகளில் சிலவற்றை பின்வருமாறு பட்டியலிடலாம்.

1. ஒரு மொழி/பன்மொழி மின்-அகராதிகள் [Mono/ Multilingual Dictionaries]
2. சொற்களஞ்சியம்[E-Thesaurus]
3. பெருந்தரவு[Corpus]
4. சொல்வலை[Word Net]
5. சொல்வங்கி[Word/Data Base]
6. மொழித் தரவுதளம்[Data Base]
7. கிளைப்பட வங்கி [Tree Bank]
8. கணிப்பொறி பயன்பாடுகளுக்கேற்ற மொழி இலக்கண விதிகள்
9. நிறுத்த சொல் பட்டியல்[List of Stop Words]. (ஆங்கிலத்திலுள்ள Smart List போன்று)
10. அறிவுத்தளத்தை உருவாக்குவதற்கான Ontology [for Building Conceptual Hierarchies]
11. சொல்வடிவங்களிலிருந்து [Word Forms] அடிச்சொல்லைக் கண்டறிவதற்கான வழிமுறைகள் [ஆங்கிலத்திற்கான Porters Stemming Algorithm போன்று]

கணிப்பொறியில் மொழிகளுக்கான அடிப்படை வசதிகள் மட்டுமல்லாது நாள்தோறும் மலையெனக் குவியும் இணையத்தகவல்களிலிருந்து தேவையான தகவல்களைக் கண்டறியவும், அக்கிடங்கில் மறைந்து ஒளிந்து கிடக்கும் ஆற்றல்மிகு அறிவுக்கருவூலங்களை மீட்டெடுப்பதற்கும்[Data Mining] மொழித்திறன் மிக்க மென்பொருள்கள் தேவைப்படுகின்றன. இத்தகைய மென்பொருள்களைத் தமிழுக்கென உருவாக்குவதற்குத் தொழில்நுட்ப வாய்ப்புகளும் சாத்தியங்களும் உள்ளன. உலக மொழிகளுக்கு பயன்படுத்தியுள்ள உத்திகள்

மற்றும் தொழில்நுட்பங்களை மாதிரியாகக் கொண்டு தமிழுக்கான மென்பொருள்களை வடிவமைக்க முன்வரும் கணிப்பொறியாளர்களின் முதன்மையான தேவை மொழிக்கருவிகளாகும். ஆங்கிலத்திற்காக ஆயத்த நிலையிலுள்ள கணிப்பொறி மொழிக்கருவிகளுடன் தமிழுக்கானவைகளை ஒப்பிடுவதற்கு துணிய இயலாது. எனவே, கணிப்பொறித் தமிழுக்கு தரமான மொழிக் கருவிகளின் உருவாக்கமும், மொழிச்சிக்கல்களுக்கு ஆற்றல்மிகு தீர்வுகளும் அவசியமாகும். இது பற்றி விவாதிக்கும் பொருட்டு பெருந்தரவைத் [Corpus] தரப்படுத்துவதற்கான முறைமைகள் எடுத்துரைக்கப்பட்டுள்ளன.

4.0 பெருந்தரவு (Corpus)

ஒரு மொழியிலமைந்த பல்வேறு வகையான உரைகளின் தொகுப்பு பெருந்தரவு [Corpus] எனப்படும். இந்தத் தொகுப்பு ஒரு மொழியின் பல்வேறு பரிமாணங்களை எதிரொலிப்பதாகவும் பலதரப்பட்ட புலங்களுக்கு [Fields] பிரதிநிதித்துவம் அளிப்பதாகவும் இருக்க வேண்டும். இத்தொகுப்பு அறிவியல் முறைப்படிபும் அமைய வேண்டும். தொகுக்கப்பெறும் உரைகள் பின்வருமாறு தெரிவு செய்யப்படுகின்றன.

- பல்வேறு ஆசிரியர்கள்/படைப்பாளிகளின் படைப்புகள்.
- கலை, மொழி, இலக்கியம், அறிவியல், தொழில்நுட்பம், மருத்துவம் போன்ற பலதரப்பட்ட பொருண்மைகளில் அமைந்த உரைகள்.
- பல்வேறு காலக்கட்டங்களில் உருவாக்கப்பட்ட உரைகள்.
- குழந்தைகள், சிறியவர்கள், பெரியவர்கள் எனப் பல வயதினருக்காக உருவாக்கப்பட்ட உரைகள்.
- முறைசார்ந்த மற்றும் முறைசாரா [Formal and Informal] உரைகள்.
- பல்வேறு பயிற்சிகள், படிப்புகளுக்கென உருவாக்கப்பட்ட பாடங்கள்.
- செய்தித்தாள்கள், இதழ்கள் போன்று பொது மக்களுக்காக உருவாக்கப்பட்ட உரைகள்.

இவை போன்று பல்வேறு வகையான உரைகளை தெரிவு செய்து முறைப்படி சரியான விகிதத்தில் தொகுக்கப்பட வேண்டும். முன்னர், பெருந்தரவுகள் கையினால் தொகுக்கப்பட்டன. தற்போது, கணிப்பொறியில் தொகுக்கப்படுகின்றன. அகராதியைப் [Dictionary] போன்றே பெருந்தரவுகளும் ஒருமொழி பெருந்தரவு [Monolingual corpus], பன்மொழி பெருந்தரவு [Multilingual corpus] என வகைப்படுத்தப்படுகின்றன. சில நேரங்களில் ஒரு மொழிக்கெனப் பெருந்தரவை உருவாக்கும்பொழுது, அவ்வரைகளை பிறமொழிகளில் மொழிமாற்றம் செய்து ஒருங்கிணைந்த பன்மொழி பெருந்தரவாக [Aligned Parallel Corpora] உருப்படுத்தப்படுகிறது. இந்தியாவில் இந்தியை மையமாகக் கொண்டு உருவாக்கப்படும் பெருந்தரவு; பிறமொழி அறிஞர்களிடம் ஒப்படைக்கப்பட்டு ஒன்றிணைந்த பன்மொழி பெருந்தரவாக ஆக்கப்படுகிறது. இம்முயற்சிகள் ஆய்வுத்திட்டளவில் முழுமைப்பெற்றதாகக் கருதப்பெற்றாலும், பயன்பாட்டளவில் அந்தந்த மொழிகளுக்கான பெருந்தரவாகக் கருத இடம் தருமா என்பது ஐயத்திற்குரியது.

4.1 பெருந்தரவும் மொழித்தொழில்நுட்பச் சோதனைகளும்

பெருந்தரவைப் பயன்படுத்துவதற்கு முன்னர் அது சரியான முறையில் தொகுக்கப்பட்டுள்ளதா எனக் கண்டறிவதற்குச் சில புள்ளியியல் சோதனைகளை மேற்கொள்ள வேண்டும். இச்சோதனைகளுக்குப் பின்வரும் புள்ளி விவரங்களைக் கண்டறிவது அவசியமாகும்.

- ஒரு பெருந்தரவில் உள்ள உரைகள்/ஆவணங்களின் எண்ணிக்கை.
- பெருந்தரவிலுள்ள சொற்களின் எண்ணிக்கை.
- பெருந்தரவிலுள்ள உரை/ஆவண நீளத்தின் தரவிலகல் [Standard Deviation] எண்.

- ஒரு உரை/ஆவணத்திலுள்ள வெவ்வேறு சொற்களின் சராசரி எண்ணிக்கை.
- ஒரு உரை/ஆவணத்திலுள்ள வெவ்வேறு சொற்களின் தர விலகல் எண்.
- பெருந்தரவிலுள்ள வெவ்வேறு சொற்களின் [Distinct Words] எண்ணிக்கை. (பெருந்தரவில் ஒரு சொல் எத்தனை தடவை இடம் பெற்றாலும் ஒரு முறை மட்டும் கணக்கில் கொள்ளப்பெறும்.)
- பெருந்தரவிலுள்ள உரை/ஆவணங்களின் சராசரி அளவு அல்லது நீளம் [சொற்களின் எண்ணிக்கையில்].

இப்புள்ளி விவரங்களைக் கொண்டு, மொழித்தொழில் நுட்பத்தில் பின்வரும் இரு புள்ளியியல் சோதனைகள் மேற்கொள்ளப்படுகின்றன.

1. ஜிப்பின் விதி [Zipf's Law]
2. டி.டி.ஆர் [TTR] Type-to-Token Ratio

4.1.1 ஜிப்பின் விதி

சரியான முறையில் தொகுக்கப்பட்டிருக்கும் தரவு தளத்தில் சொற்களின் நிகழ்வு எண்ணிக்கையையும் [Frequency(f)], தரவரிசை எண்ணையும் [Rank(r)] பெருக்கிவரும் தொகை [$f \times r$] ஒரு மாறிலி [Constant] ஆகும். இதுவே ஜிப்பின் விதியாகும்.

பெருந்தரவு தொகுக்கப்பட்டுள்ள தன்மையைக் கண்டறிய அதன் மீது ஜிப்பின் விதியைச் சோதித்துப்பார்க்க வேண்டும். இதற்கு பெருந்தரவிலுள்ள சொற்களைக்கொண்டு சொல் எண்ணிக்கை அட்டவணை [Word Frequency Table] ஒன்றை உருவாக்க வேண்டும். இந்த அட்டவணையின் படிவம் பின்வருமாறு அமையும்.

வ.எண்	சொல்	சொல்லின் எண்ணிக்கை [Frequency(f)]	நிகழ்வு	தரவரிசை எண் [Rank (r)]	நிகழ்வு எண்ணிக்கை $f \times r$ தரவரிசை எண் $f \times r$
-------	------	-----------------------------------	---------	--------------------------	---

இந்த அட்டவணையில், சொற்கள் நிகழ்வெண்ணிக்கையின் அடிப்படையில் இறங்கு வரிசையில் தொகுக்கப்பட்டிருக்க வேண்டும். அதிக எண்ணிக்கையில் இடம் பெற்ற சொல் முதலாவதாகவும், அதற்கு அடுத்த சொல் இரண்டாவதாகவும் அமையும்.

மேற்கண்ட அட்டவணையில் $f \times r$ எல்லா சொற்களுக்கும் ஒன்றாகவே அமைந்தால் பெருந்தரவு ஜிப்பின் விதிக்குப் பொருந்துகிறது என்று பொருள். அறிவியல் முறைப்படியும் சரியாகவும் பெருந்தரவு தொகுக்கப்பட்டிருப்பதாகக் கொள்ளலாம். அவ்வாறு இல்லையெனில், பெருந்தரவில் மேலும் சில உரைகளைச் சேர்த்து மேம்படுத்த வேண்டும்.

4.1.2 டி.டி.ஆர் மதிப்பு

டி.டி.ஆர் மதிப்பை பின்வரும் சூத்திரத்தின் மூலம் கண்டறியலாம்.

டி.டி.ஆர் (TTR) = பெருந்தரவிலுள்ள வெவ்வேறு சொற்களின் (Distinct terms) எண்ணிக்கை / மொத்த சொற்களின் எண்ணிக்கை.

டி.டி.ஆரின் மதிப்பு பூஜ்ஜியத்திலிருந்து ஒன்றுக்குள் இருக்கும் ($0 \leq TTR \leq 1$). டி.டி.ஆரின் மதிப்பு பூஜ்ஜியத்தை நோக்கி அமைந்திருந்தால் (அதாவது 0.5க்கு குறைவாக இருந்தால்) பெருந்தரவில் வெவ்வேறு சொற்கள் அதிகமாக இடம் பெற்றுள்ளன எனப்பொருள்படும். டி.டி.ஆரின் மதிப்பு ஒன்றை நெருங்கி

இருந்தால் (அதாவது 0.5 மற்றும் அதற்கு மேலிருந்தால்) பெருந்தரவில் வெவ்வேறு சொற்கள் குறைவாக உள்ளன என்று பொருள்.

வேறு வகையாகக் கூறின், பெருந்தரவு குறைந்த ஆவணங்கள்/ உரைகள் மூலம் தொகுக்கப்பட்டிருந்தால் டி.டி.ஆர் மதிப்பு ஒன்றை நோக்கிச் செல்வதற்கு வாய்ப்புள்ளது. எனவே, பெருந்தரவில் பல்வேறு வகையான உரைகள் மேலும் சேர்க்கப்பட வேண்டும். இவ்வாறாக, டி.டி.ஆர் மதிப்பு பெருந்தரவைச் சரியாகத் தொகுக்க உதவுகிறது.

4.1.3 பெருந்தரவின் தர நிர்ணயம்

தரமான பெருந்தரவை உருவாக்க கருத்தியல் [Theory] அளவில் இத்தகைய விதிமுறைகள்/வழிமுறைகள் உள்ளன. இவைகள் நடைமுறையில் எந்த அளவு பின்பற்றப்படுகின்றனவோ அந்தளவுதான் உருவாக்கப்படும் பெருந்தரவின் தரம் அமையும். உருவாக்கப்பட்ட பெருந்தரவு பலரின் பயன்பாட்டிற்கு வந்தால்தான் பயன்பாட்டாளர்களின் மதிப்பீடுகளுக்கு உட்பட்டு செம்மை பெறும். பொதுவாக, பெருந்தரவை ஒரு மொழியின் முழுமைப்பெற்ற அறிவுக்கான மாதிரியாகக் கருதாமல், ஒரு மொழியின் உண்மை நிலையைப் பற்றி விவாதிப்பதற்கான ஆய்வுக்களமாக மட்டுமே கருத வேண்டுமென்ற கருத்து மொழி வல்லுநர்களின் ஒரு சாராரிடம் உள்ளது.

மொழிக்கருவிகளை வெளியிடும்பொழுது வடிவமைப்பின் நோக்கங்கள், வரையறைகள், உருவாக்கப்பட்ட முறைகள், தரப்படுத்தியமைக்கு மேற்கொண்ட சோதனைகள் போன்ற விவரங்களைத் தருதல் நலம். இது நம்பிக்கைத் தன்மையைத் தரும். பின்னர், பயனாளிகளிடமிருந்து நடைமுறைப் பயன்பாட்டளவு, பிழைகள், மதிப்பீடு, மேம்பாடு செய்வதற்கான கருத்துக்கள் போன்ற எதிரொலிகளைப் பெறுவது அவசியமாகும். மென்பொருள் வடிவமைப்பின் இவ்வாறான முறைகளை முறைப்படி பின்பற்றியதால்தான், 50 ஆண்டுகளுக்கு முன்னர் தொடக்க நிலையில் மின்னணு தட்டச்சுப் பொறியாக கணிப்பொறியை இயக்க உதவிய சொல்லாய்வி [Word Processor] கற்பனைக்கும் எட்டாத வளர்ச்சிகளைத் தற்போது பெற்று விளங்குகிறது. இதுபோன்றே உலக மென்பொருள்கள் பலவும் அவ்வப்பொழுது வளர்ச்சி பெற்று மேம்பாடு அடைந்து வருபவைகளேயாகும்.

4.1.4 தரமான பெருந்தரவின் ஆக்கப் பயன்கள்

செம்மையும் சீர்மையும் பெற்ற பெருந்தரவுதான் தரவிலிருந்து மொழிக் கருத்தியலுக்கானப் [Theory] பாதையை வகுக்கும். மேலும் மொழிக் கொள்கைகளைச் சோதனை செய்வதற்கும், மொழியில் புதிய விதிகளை உருவாக்குவதற்கும், பழைய விதிகளை மதிப்பீடு செய்வதற்கும், புள்ளி விவரப்பகுப்பாய்வு செய்வதற்கும் மொழியைக் கற்கவும் கற்பிக்கவும் அகராதி போன்ற மொழிக்கருவிகளை உருவாக்குவதற்கும் உறுதுணையாக அமையும். இவ்வாறு உருவாக்கப்படும் பெருந்தரவு ஆய்வு நிலைகளில் சோதனைகளுக்கு பயன்படும் வகையில் கருத்துவிளக்க [Annotation] செயலாக்கம் பெற்றதாகவும் அமைய வேண்டும். எடுத்துக்காட்டாக, சொல்லின் இலக்கணக்கூறுகளை அடையாளப்படுத்தும் POS Tagging, அடிச்சொல்லை [Base] வெளிப்படுத்தும் Tags, சொற்றொடர் மற்றும் வாக்கியத்தின் கட்டமைப்பை வெளிப்படுத்தும் Structural Level Tags ஆகியன பெருந்தரவின் தரத்தை மேம்படுத்தும்; பயன்பாட்டை அதிகரிக்கும். மொழி வல்லுநர்களைத் தவிர பிற துறை அறிஞர்கள் குறிப்பாக கணிப்பொறியாளர்கள் மொழிப்பயன்பாடுகளுக்கென உருவாக்கும் மென்பொருளின் வடிவமைப்பிற்கு உதவும். தகவல் தொழில் நுட்பத்தில் உலக மொழிகளைக் கையாள்வதற்கு வடிவமைக்கப்பட்டுள்ள மென்பொருள்கள் போன்று அந்தந்த மொழிகளுக்கான மென்பொருள்களை வடிவமைக்க இக்கருவிகள் உதவும். குறிப்பிட்ட பயன்பாட்டிற்கென ஒரு மொழியில் உருவாக்கப்படும் மென்பொருள் எந்தளவுக்கு பயன்படும் என்பதைச் சோதித்தறிய இயலும். மொழிக்கான மென்பொருள்களின் செயல்திறனை, பயன்பாட்டுத்திறனை, துல்லியத்தைச் சோதனையிட பெருந்தரவு ஒரு ஆய்வுத்தரவாகவும் சோதனைக் களமாகவும் விளங்கும்.

சுருங்கக் கூறின், மொழித் தொழில் நுட்பம் [Language Technology], மொழி பொறியியல் [Language Engineering] போன்ற பயன்பாடுகளுக்கும், தகவல் தொழில்நுட்பக்கருவிகளில் அன்றாட பயன்பாடுகளுக்கான மென்பொருளின் தர மதிப்பைச் சோதித்தறிவதற்கும் நல்ல மாதிரியாக [Model] அமையும்.

5.0 ஒலி பெயர்ப்பு [Transliteration]

தமிழ்ச் செய்திகளைக் கணிப்பொறியில் உள்ளீடு செய்யும் முறைகளுள் ரோமன் வரிவடிவத்தைப் பயன்படுத்துவதும் ஒன்று. ஒரு மொழியின் சொல்லை அதன் ஒலிப்பு மாறாமல் பிற மொழியின் எழுத்துக்களில் மாற்றம் செய்தல் 'ஒலி பெயர்ப்பு' அல்லது 'எழுத்துப் பெயர்ப்பு' எனப்படும். ஒலி பெயர்ப்பு செய்யும் பணிகள் நூலகங்களில் சில நூற்றாண்டுகளுக்கு முன்னரே தொடங்கப் பெற்றுள்ளது. ஒலி பெயர்ப்பில் எழுத்தின் ஒலி அளவை [மாத்திரையைக்] குறிக்க ஆங்கில எழுத்துக்களின் மேல் அல்லது கீழ் புள்ளி [.,], சிறிய கோடு, பிறை, போன்ற சிறப்புக் குறியீடுகளைப் பயன்படுத்தும் முறைகளும் [எ.கா.tamil] ஆங்கில எழுத்துக்களை மட்டுமே பயன்படுத்தும் முறைகளும் [எ.கா.tamizh] நடைமுறையில் உள்ளன. ஒலி பெயர்ப்பு முறைகள் 1888 ஆண்டில் தரப்படுத்தப்பட்டிருக்கின்றன. இணையத்தில் தமிழிலுள்ள தகவல்களைத் தேட "கூகுள்" [Google] தேடும் பொறியில் தமிழ்ச் சொற்களை உள்ளீடு செய்ய ஒலிபெயர்ப்பு முறையைக் கையாள வாய்ப்பு அளிக்கப்பட்டுள்ளது. இதிலுள்ள முறையும் தமிழ் மென்பொருள் பிறவற்றில் பின்பற்றப்படும் முறைகளும் வெவ்வேறானவை. எனவே, தற்போதைய கணிப்பொறி தொழில் நுட்பத்திற்கேற்ப ஒலிபெயர்ப்பில் உலகத்தரம் வாய்ந்த முறைமைகளைத் தெரிவு செய்ய வேண்டியதும் தரப்படுத்த வேண்டியதும் மேன்மைக்குரியன.

6.0 நிறைவுரை

பெருந்தரவிற்கானதைப் போன்றே ஒவ்வொரு மொழிக்கருவிகளை உருவாக்குவதற்கும் தரப்படுத்துவதற்கும் உலக மொழிகளில் பல முறைமைகள் பயன்படுத்தப்படுகின்றன. கட்டுரையின் பக்க வரையறையைக் கருத்தில் கொண்டு பெருந்தரவு மட்டும் எடுத்துரைக்கப்பட்டுள்ளது. சுமார் 60 நாடுகளில் புலம் பெயர்ந்து வாழும் தமிழர்களால் இணையத்தமிழ் பிறந்தது. தரணி முழுதும் தமிழ்ப் பரப்பும் சமுதாயத் தொடர்புச் சாதனமாக இணையம் பயன்படுகிறது. தமிழ் இலக்கியங்கள், கல்வெட்டுகள், சுவடிகள் போன்றவற்றைப் பாதுகாத்துப்பேணுதல், புதிய தொழில் நுட்பங்களைக் கையாண்டு மொழியை வளப்படுத்துதல் போன்ற நோக்கங்களோடு எண்ணற்ற தமிழ்த்தளங்கள் வந்தவாறே உள்ளன. உலகத் தமிழர்களை ஒன்றிணைக்கும் பாலமாகத் தமிழ் இணைய தளங்கள் விளங்குகின்றன. மேலும், அரசுக்கும் குடிமக்களுக்கும் [Government to Citizen G2C] உள்ள உறவைப் பலப்படுத்துவதற்கு தகவல் பரப்பு ஊடகமான இணையத்தில் மொழிகளின் பயன்பாடுகளைக் கொணர மைய மற்றும் மாநில அரசுகள் முனைந்துள்ளன. இந்நிலையில், கணிப்பொறித் தமிழ் வளர்ச்சிக்காக உலகத்தமிழர்கள் கையில் எடுத்துள்ள முயற்சிகள், முன்னேற்றங்கள், ஒன்றிணைப்புகள் ஆகியன மொழி, மொழியியல், கணிப்பொறியியல் எல்லாவற்றையும் ஒன்று திரட்டிய பரந்த பார்வையைத் தரும். ஒரு துறையினர் மட்டும் தெரிந்து கொள்ள முடியாத, உணர்ந்துகொள்ள முடியாத புதிய கருத்துக்களைப் புலப்படுத்தும். செம்மொழி திட்டங்கள் தர உள்ள வாய்ப்புகளைப் பெற உதவும். இதனால் நல்ல பயன்கள் விளையும்.

நூற்பட்டியல் [Bibliography]

1. பாஸ்கரன்.ச., தமிழில் கணிப்பொறியியல்-கணிப்பொறியில் தமிழ், 2003, உமா பதிப்பகம், தஞ்சாவூர்.
2. இராசேந்திரன்.ச. , பாஸ்கரன்.ச. , தமிழ் மின் சொற்களஞ்சியம், 2006, தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்.
3. Baskaran S., Content Based E-Mail Classification System by Applying Conceptual Maps, 2009, IEEE.

Automated Processing of Census Forms in Tamil

Shashi Kiran¹, Rituraj¹, Suresh Sundaram¹, Swapnil Belhe², AG Ramakrishnan¹

MILE Lab, Dept of Electrical Engineering,
Indian Institute of Science, Bangalore 560 012.

Center for Development of Advanced Computing, GIST Group, Pune
shashi.reach@gmail.com, sureshsundaram1980@gmail.com,
swapnilb@cdac.in, ramkiag@ee.iisc.ernet.in

1. Abstract

This paper describes automatic form filling system for collecting census data based on online handwritten character recognition for Tamil. The aim is to facilitate easy digitization of Indian language ink-data gathered from field. The application interface is designed in such a way that the same application can be adopted for other Indian languages. It also describes the common interface framework required to facilitate this multiple language recognition engines. For Tamil or any other Indian scripts; inputting isolated symbols is not practical, hence the application uses non-isolated character recognition. The application incorporates simplified method of form design, layout analysis, engine error correction; engine level limited vocabulary based post-processor, validations etc. The performance evaluation of this application is also carried out against the traditional methods and promising results are obtained. The same system can be easily adopted for other types of forms required to be filled in Tamil e.g forms used by Government institutions, Banks etc.

2. Introduction

India is one of the very few countries in the world, which has the proud history of holding census every ten years uninterruptedly since 1872. The census provides information on size, distribution and socio-economic, demographic and other characteristics of the country's population. The data collected through the census are used for administration, planning and policy making as well as management and evaluation of various programs by the government, NGOs, researchers, commercial and private enterprises, etc. [1]

Government agents visit each and every household in the country and collect complete data about the people and the condition of their houses. These agents gather this data by filling printed forms. The filled forms are then used to manually enter data in computers. This procedure takes long time and is prone to human errors while re-entering data manually. The data collection and the data storage are the two basic stages of this activity.

This census form processing system requires the members to carry the digital pads to the site, collect the information on this device, once the data collection is done plug in this digital pad to the computer and this application will generate the database.

Gathering information by using printed forms is predominantly used in governmental, educational, banking domain. Even on-field surveys are carried out using local languages. Traditionally, the major surveys like census are conducted on field by pen & paper wherein a printed form is filled by a surveyor. The surveyor collects information on the printed forms; and when the data collection is completed, forms are first scanned and then send for the verification and data entry. This method works well for small surveys requiring limited information. But when the survey requiring detailed information is to be carried out across cities, states and country; the traditional method becomes time consuming. The large amount of time is spent on scanning, verification and data entry.

The goal of this automated census form filling system is to provide simplified form filling process which works with very low cost digitizers and considers major Indian languages. Reduction in cost of digitization is also considered during design of this system.

In this system, the data collection process remains the same except an offline digitizing tablet is used for writing forms. The paper is kept on the tablet and the user writes on the paper by using special pen. The currently available offline digitizing tablets provide the advantage of retaining hard-copy of the filled forms for future use. These tablets with no display capture the ink data as X-Y co-ordinates and pen pressure. This collected ink data in X-Y co-ordinate format is then submitted to the form processing application which converts it to computer editable text.

3. Data Collection

The data collection phase is most crucial phase of the form processing application. The procedure for collecting data using digital tablet is similar to what currently being followed by data collectors. Instead of only-paper based forms the battery backed digital tablet is used along with digital ink based pen as shown below.



The data collected on such digital device can then be downloaded on the Computer for further processing. This is where Tamil online handwritten engine come into effect.

The Surveyor who is collecting the data is required to follow the manifest for collecting data which is similar to traditional data collection. We have collected data from ten writers of Tamil for testing by using G-Note 7000 digitizing tablet working at 160 points/sec.

4. Architecture of System

This system is primarily divided into four modules for simplicity during development, the modules are,

- User Interface
- Input Data Extraction module
- Recognition Engine Interface module
- Database Generation module

The application architecture is kept broad to easily plug-in different components. The application is currently designed to support two different types of digitizers namely Genius G-Note and iBall TakeNote but can easily be adopted to other devices. The Genius device produces the files with TOP extension while iBall produces files with DHW extension. As explained in the previous section the raw ink data (x, y co-ordinates) coming from the device is given to the application as input. Figure 1 explains the broad architecture of the application.

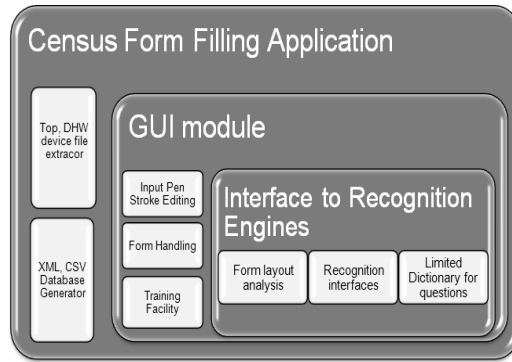


Figure 1: Architecture of Tamil Form Processing Application

The input data is displayed in full page format for verification. If the data contains any abnormal and misaligned ink strokes, they are removed by the operator. The verified data is then fed to the layout analysis module. This module aligns the input data to the pre-defined template using positional analysis. This way, the answers by the client are mapped to corresponding questions. Once the correct alignment is achieved the text/numeral fields are separated from ticks. The text and numerals are given to Indic engines for recognition. This is achieved by using common interface framework used to link the Tamil handwritten recognition engine to the application. The output of the engine contains the recognized text with confidence scores. The confidence score is used to provide the operator with choice to select the correct outcome. If the score is high enough then no selection choice is provided.

The operator is also provided with easy Tamil text editing functionality like virtual keyboard.

5. Common Interface Framework

Since this form processing system is designed to accommodate all Indian languages along with Tamil, there are many recognition engines to be interfaced to the system; hence there is a great need to have a standard communication protocol between different handwritten recognition engines and census application. A common Interface framework provides this vital link between the recognition engines and the application. This framework provides the flexibility to add/remove recognition engines. Also an entire new application can be built around this framework without the need of knowing the intricacies of online character recognition technologies. So this framework provides the scalability of adding new engines to the existing application. Simply put, it act as messenger between the application and the engine and thus helps to change any layer on the fly (i.e. we can change the application without disturbing the rest of setup or vice versa).

6. Pre-processing

6.1 Automated Segmentation

The segmentation of the validated data reduces the overhead of the Tamil recognition engine and helps in fast and accurate recognition of the handwritten text. All strokes in a page are given to segmentation module. In this module, all the strokes are arranged as per their position in a page irrespective of their order of occurrence. This module segments the lines based on horizontal projection of the strokes. Once the lines are segmented, the gaps between histogram (vertical projection) of each stroke on the line are clustered into within-word-gaps (WWG) and between-word-gaps (BWG) [2]. The words are separated by WWG.

6.2 Form Cleaning

The digital ink data collected by the surveyor often has noisy, unclean ink. Sometimes, the surveyor himself writes some notes, annotations, comments on the pages inside or outside the page boundaries. There also could be scrubbing which is large enough, crossing the full page or questions by digital ink. There could be overwriting, page misalignment or completely missing the pages of the form. This kind of data is required to be cleaned before proceeding for recognition. Otherwise the recognition accuracies could be very poor. Since it is difficult to clean all such noise automatically, the form cleaning is done semi-automatically. The cleaned, verified and segmented data is passed for recognition by the language specific engines.

7 Additional Features

7.1 Tick detection

The forms used for census survey contains various data fields like text, numerals, check boxes etc. The surveyor is supposed to tick into the boxes wherever necessary. First, the check boxes needs to be separated from other fields. This is achieved by positional analysis of the form. Once the relative position of the ticks is identified, it is required to separate actual tick marks from the scrubbing and other unwanted strokes. The tick detection is used for detecting the check mark on the form; this includes fields like radio buttons or check boxes. There are various ways of recognizing ticks in this implementation we have used Dynamic time warping (DTW). It is an elastic matching algorithm for matching the similarity between two given sequences. The similarity matching is based on the distance measure. The sequence with minimal distance is considered for optimal match. Since the distance based sequence matching is not feasible for every point in sequences we need to put two conditions i.e. boundary and continuity condition. The continuity condition decides how much the matching is allowed to differ from linear matching [3, 4]. The Boundary condition states that first and last points of sequences will be matched with each other. The continuity condition decides the measure of elasticity given by the formula:

$$\frac{N_2}{N_1}i - cN_2 \leq j \leq \frac{N_2}{N_1}i + cN_2$$

Where c is the continuity constant, N_1 and N_2 are the number of points in first and second curve respectively. The points i and j of the first and second curve respectively can be matched only if above condition is satisfied. For $c=0$ the resulting match is same as linear matching.

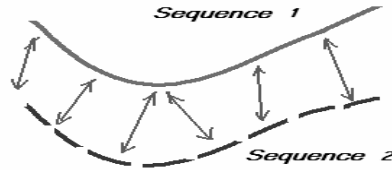


Figure 2 DTW matching of two ticks

DTW is trained on data collected from 20 clients, for ten different styles of writing ticks. Certain conditions are applied to the data before passing the data to DTW such as checking for number of strokes.

7.2 Error Correction Measures

The recognition accuracy of the underlying Tamil handwritten recognizer is not 100% hence it becomes responsibility of the application developer to highlight those inaccuracies to the *operator* in a mild way. The ideal application for such kind of systems would be one which not only gives the operator the helping hand in case of a recognition error but also trains the engine over time by understanding corrections made by operator. It also allows editing the input file with stroke addition/deletion functionality and saving the same.

Significant percentages of operators are bound to get irritated in case of repeated errors on part of recognition (governed by recognition engines). An effective user interface in this case would be able to minimize the edits required by use of automatic focusing, multimodal inputting, easy suggestion list etc.

The application is built with stroke correction facility. If the ink data contains spelling mistakes or ambiguities then the operator can re-write the strokes and save it for future use. The operator is also given the facility to type using virtual keyboards. The complete application is localized for Tamil.

Since each field is separated by the values it may have, like text, numerals, numerals with symbols etc. The application performs the broad validity check on each output returned by the engine. This helps in narrowing down the errors for each field.

Based on the probability measure returned by the Tamil recognizer engine, the application takes the decision on whether the recognized text is suitable for displaying. If not, the field is marked with red color pointing to the operator about possible error that may require relook.

7.3 Standardized output format

The operator is given the flexibility to save the recognized data in Unicode based Comma Separated Variable (CSV) files or Extended Mark-up (XML) files. These files can be easily imported by any of the database engines like Oracle, SQL etc.

7.4 Wordlist for Post-Processing

The common interface framework includes the interface to send the domain specific Tamil wordlists to the handwritten recognizer. The application developer changes the wordlists as per different requirements of the applications and sends it to engines so that language models can be applied in order to improve the performance. The structure of the wordlist is defined by the framework.

8. Results and Discussions

We compared the performance of this automated census form filling application with the traditional ways of conducting census surveys. The performance of the application is highly dependent on the recognition accuracy of the Tamil Online handwritten recognizer. Application uses some domain specific knowhow of census form data to improve on the base recognizer performance. In the traditional surveys, the filled forms are scanned and then manual data entry is done. In this experiment the census form used for evaluation was very comprehensive spanning over 10 pages. Following table shows the composition of the form used for the survey.

Field Type	Number of fields per form
Text (with wordlist)	8
Text (without wordlist)	17
Numerals	22
Check boxes (multiple ticks)	10
Radio boxes (single ticks)	27

The text fields like clients education background, spoken language, city, district, state etc. were backed with limited dictionary while fields like name, last name, address etc. were without dictionary.

For evaluating performance of traditional census data entry, the forms were given to 3 data entry operators who were regulars in Tamil typing. Total data was collected from 10 writers; each form contained 10 pages. The same forms are processed through this application. Following table shows the comparative results of traditional manual data entry and online character recognition based application.

	Average Time (Time/Form) (10 pages per form)	Average Error in output
Manual Data Entry	89.09 sec.	2.68 %
Automated Form processing + Verification	30.37 sec.	13.88 %

As seen from the above table, the manual data entry took more time but numbers of errors were less. Note that the page scanning time required for manual data entry is not considered in the above table.

9. Conclusion & Future Work

In our experiment, the online handwritten character recognition based form processing for Tamil clearly showed a promising area for further research especially where collecting huge quantities of data from field and converting it into editable text is concerned. In this application we retain the paper based form filling & editing thus allowing more natural text inputting and reduce cost incurred on display based inputting devices. In future, we would like to study the impact of the Indian language word models on user acceptance of online handwriting recognition.

10. Acknowledgment

The authors would like to thank Consortium for "Online Handwritten Character Recognition" and Technology Development for Indian Languages (TDIL), Department of Information Technology (DIT), Government of India for funding this consortium project. The authors would also like to thank all the consortium chief investigators and members consisting of IISc-Bangalore, ISI-Kolkata, IIT-Madras, IIIT-Hyderabad, CDAC-Pune for their valuable inputs.

References

1. Ashish Krishna, Girish Prabhu, Kalika Bali, Sriganesh Madhvanath, "Indic scripts based online form filling - A usability exploration", 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, 2005
2. Soo H. Kim, S. Jeong, Guee-Sang Lee, Ching Y. Suen, "Word Segmentation in Handwritten Korean Text Lines Based on Gap Clustering Techniques", Proceedings of the sixth International Conference on Document Analysis and Recognition (ICDAR-'01), Seattle, WA, pp. 189-193
3. Ralph, Niels and Louis, Vuurpijl, "Dynamic Time Warping Applied to Tamil Character Recognition," Proceedings of Eight International Conference on Document Analysis and Recognition (ICDAR'05), pp. 730-734
4. N. Joshi, G. Sita, A. G. Ramakrishnan, and S. Madhvanath, "Comparison of elastic matching algorithms for online Tamil handwritten character recognition," Proceedings of the Ninth International Workshop on Frontiers of Handwritten Recognition (IWFHR'04), pp. 444-449
5. XStroke: Full-screen Gesture Recognition for X, Carl D. - Worth Information Sciences Institute University of Southern California Arlington.
6. UPX- The best from UNIPEN and ink ML, - <http://unipen.nici.kun.nl/upx/>, 2002.
7. Swapnil Belhe, Srinivasa Chakravarthy, A.G. Ramakrishnan, "XML Standard for Indic Handwritten Indic Database" Proceedings of International Workshop on Multilingual OCR (MOCR-09), Barcelona, Spain, July 2009.

Pattern based English-Tamil Machine Translation

S. Saravanan
sarwanster@gmail.com

Dr.A.G. Menon
menon.govindankutty@gmail.com

Dr. K. Soman
kp_soman@amrita.edu

Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore

Introduction

The native languages all over the world are growing rapidly along with the growth of technology, in general, and information technology, in particular. On the one hand the world experiences a growth in the native language and on the other hand precious and nascent information come through foreign languages. Literacy in the mother tongue is no longer enough to follow the information supplied by the other languages. Because of this ever increasing gap and the speed with which information are supplied, it is necessary to bridge this gap with the help of modern technologies as early as possible. It is in this context that we are working on a Machine Translation (MT) system.

Even though, there are several approaches to develop a MT system, LTAG-based MT and SMT are very prominent. SMT is far away from tasting the success in case of agglutinative languages like Tamil. The only available MT system for English-Tamil is LTAG-based developed by AMRITA in collaboration with CDAC, funded by DIT. Transfer rules in LTAG-based MT are a pair of trees. Writing a new transfer rule is not very easy.

This paper proposes the use of pattern-based reordering rules in MT. Unlike the translation patterns in pattern-based CFG for MT (Koichi Takeda), the pattern based reordering rules are not lexicalized and the features and agreements are not handled in the rule. The source rules of these reordering rules are not used for parsing the source sentence. These rules are used to reorder the source phrasal structure to get the target phrasal structure. The English words are translated and lexicalized separately and this process is called lexicalization. The features and agreements are collected from the parse tree and used for synthesizing the words.

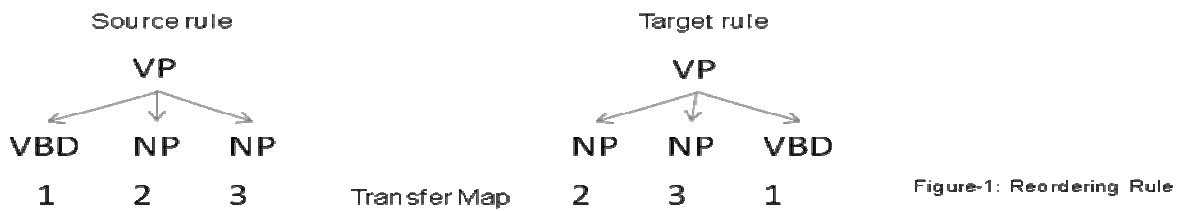
Parsing

One of the first steps is the identification of the structure with a parsing algorithm. Different parsing algorithms are available for the syntactic analysis of the source sentences. We are using the Stanford PCFG parser for the analysis of the source sentences in English. The sample parse output of the English sentence "Ram gave him a book" is shown in the figure-2.

Pattern based Reordering

A pattern is a pair of CFG rules (Takeda, Koichi: 1996). These rules give the equivalent structures in the source and target languages. On the basis of the patterns the reordering rules are formulated to facilitate the machine translation. They reflect the translation patterns of the source and target languages. For example, the following reordering rule is based on an English- Tamil pattern: VP (VBD NP NP) → VP (NP NP VBD) || 1:2 2:3 3:1

'VP (VBD NP NP)' is the English CFG rule ('VP' is the root node and 'VBD', 'NP', 'NP' are the children nodes in the tree representation) called a source rule and 'VP (NP NP VBD)' is the Tamil CFG rule called a target rule. "1:2 2:3 3:1" is the transfer link. The tree representation of the above rule is shown in the figure-1 below.



The transfer link contains the order of the children nodes of the target rule. "1:2 2:3 3:1" says first child of the target rule is from second child of the source rule; second child of the target rule is from third child of the source rule, so on and so forth.

The parse tree of the source language is checked against the source rules. If any match is found in the parse tree, then the source rule is replaced with the corresponding target rule. For example, the parse tree of the English sentence, "Ram gave him a book" is (S (NP (NNP Ram)) (VP (VBD gave) (NP (PRP him)) (NP (DT a) (NN book)))). This English phrasal structure is checked with the available reordering rules for finding a match. The pattern in the source language such as 'VP (VBD NP NP)' is present in the English phrasal structure and it is eligible for undergoing the reordering rule. The pattern "VP (VBD NP NP)" of the source language is thus replaced with its counterpart "VP (NP NP VBD)" in the target language.

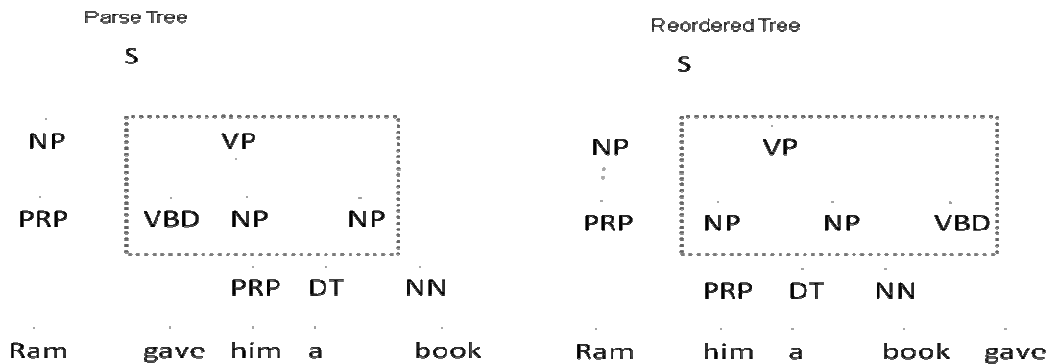


Figure-2: Transformation of Parse Tree to Reordered Tree

Tamil shows a very high degree of flexibility in ordering the words within a sentence. The position of the words can be easily transposed without much change in the meaning. For example, “Ram gave him a book” can be reordered in multiple ways in Tamil, and the most common ways are: Ram him a book gave, Ram a book him gave, Him a book Ram gave, etc,. The predicate verb takes mostly the last position. In our system, the reordering rules are strictly one to one map. Every source rule is mapped to one target rule. Based on the most common usage, the target rule is formulated. The Tamil clausal structure is more rigid and shows little flexibility. For example, “Ram, who is smart, gave him a book.” is reordered as “(smart Ram) (him) (a book) (gave)”. Here the adjectival clause ‘who is smart’ has to be positioned before the noun ‘Ram’ in Tamil.

Reordering Algorithm

Let T be the parse tree with N number of nodes that we process for reordering; R be the number of reordering rules; S_t be the source rule tree; G_t be the target rule tree; $Sc(n)$ be the sub tree of the node, n with the depth one. For example, the $Sc(VP)$ in the parser tree is VP (VBD NP NP)

Re-order (T):

```

Visit node n
If n equals root ( $S_t$ )
  For each  $S_t$  of reordering rule R
    If ( $Sc(n)$  equals  $S_t$ )
      Replace  $Sc(n)$  with target tree,  $G_t$ 
  For each child c of n
    Re-order (sub tree(c))
  
```

The Re-order algorithm with the N number of nodes in the parse tree, T and R number of reordering rules has the complexity of $O(N*R)$.

Lexicalization

The words in the terminal node of the reordered tree are lemmatized and the equivalent of that lemma in target language is replaced in the terminal node of the reordered tree. For example, lemma of the source word ‘gave’ is ‘give’ and its equivalent in target language is ‘koTu’. The target equivalents are found in the root word lexicon which contains the root forms of the source words and target words along with the POS category.

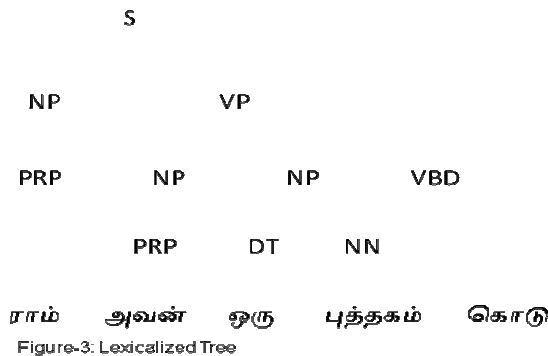


Figure-3: Lexicalized Tree

Ram	→	ராம்
gave	→	give → கொடு
him	→	அவன்
a	→	ஒரு
book	→	புத்தகம்

Figure-4: Lexicon

The major challenge in the lexicalization process is semantics. There would be multiple senses for the source word. In that case, the current system has the provision for multiple outputs. In the case of the functional words and auxiliary words, there are no direct equivalents in Tamil. For example, “The auditorium is decorated for the college day celebrations”, the words ‘for’ and ‘the’ do not have equivalent words in Tamil. ‘for’ comes as a post position (kkAka) in Tamil and is synthesized along with the Noun ‘college-day’. The auxiliary word ‘is’ has no direct equivalent in Tamil. The words ‘college day’ consists of two entities in source and the equivalents for these two words separately are ‘kallUri nAL’ which is semantically wrong. The compound nouns are to be considered as one entity. The equivalents of the words ‘college-day’ as single entity are ‘kallUri ANTU vizha’. The compound nouns are found and marked as single entity using the POS of the source sentence. The POS of this example sentence is, “The/DT auditorium/NN is/VBZ decorated/VBN for/IN the/DT college/NN day/NN celebrations/NNS”. Here the words ‘college’ and ‘day’ are annotated as ‘NN’. The consecutive words that annotated as ‘NN’ are marked as single entity in the pre-processing, and during lexicalization the equivalents of this entity are found and replaced in the terminal node.

Transliteration

Transliteration is an automatic method that converts words/characters in one alphabetical system to corresponding phonetically equivalent words/characters in another alphabetical system(Vijaya MS, 2009). The Named-Entities are transliterated from English to Tamil using the tool that based on ‘Sequence Labeling Approach’. 30k person names and place names (English-Tamil pairs) are used for training using SVM. The words that are not present in the root lexicon (Out Of Vocabulary words) are also transliterated in the same manner.

Synthesizer

The Morphological synthesizer glues the lemma and the morphemes to form a word using orthographic and morphophonemic rules. The lite version of Amrita Morphological Analyzer and Generator (AMAG) is used for synthesizing the words. The synthesizer requires the information along with lemma as an input. This information has to be gathered from the lexicalized target phrasal structure, the parse tree of the source and from the typed dependency information. The lexicalized tree has the lemma of the target language in the terminal nodes. Synthesizing these lemmas is very important in the process of translation. Synthesizing the noun gives the relationship between two nouns and in the case of verb it glues the TAM and gender information with the verb.

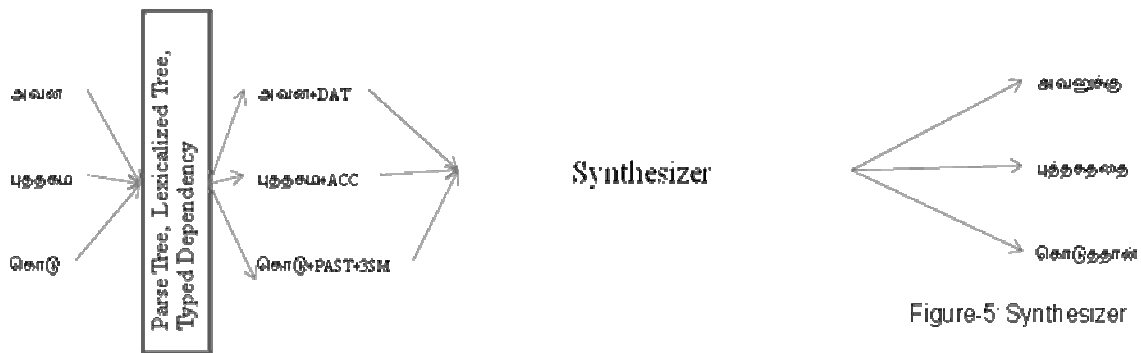


Figure-5: Synthesizer

For example, the word 'avan' is given as input to the synthesizer with the morpheme information 'DAT' as 'avan + DAT' (dative). The first level of the synthesizer replaces the morpheme information with the correct morpheme that to be synthesized with the lemma as 'avan + ku'. The second level of the synthesizer applies the spelling rule on 'avan + ku' and after the application of all possible rules; the second level outputs the synthesized word 'avanukku'.

Testing and results

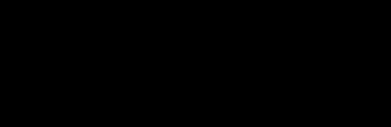
The current system is tested with the corpora of Tourism domain that we developed for DIT funded EILMT project. On testing with the corpora of the size of 2000 sentences, 60 % of the sentences are translated well and 70 % of the sentences are comprehensible. On the module wise testing, 80 % of the sentences are reordered perfectly, 60 % of the sentences are lexicalized properly with its correct target word, and the accuracy of the transliteration module is 93.3 % and more than 90 % of words that process through synthesizer module are synthesized properly, provided the information extracted and given as input to the synthesizer module is correct.

Conclusion

At present, the prototype version of the MT is developed with the 20k lexical entries and very few (40) reordering rules. In future, Word-Sense-Disambiguation-module can be plugged-in to the system to resolve semantics ambiguities. The system with all the necessary modules are in place, scalability is a key to improve its performance. Transliteration, Morph-synthesizer and extracting features are on its own a big task and these have to be enhanced as well to improve the overall performance of the system along with the root word lexicon and the reordering rules.

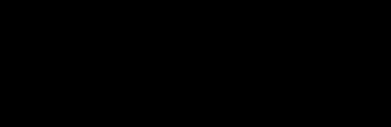
Bibliography

1. Abeilld, A., Schabes, Y. and Joshi, A.K. 1990. "Using Lexicalized tags for Machine translation." *Proceedings of the 13th International Conference on Comparative Linguistics*, August 1990, Vol. 3: 1-6.
2. Menon, A.G, Saravanan S, Loganathan R, Soman, K.P. 2009. "Amrita Morph Analyzer and Generator for Tamil: A Rule Based Approach". *8th Tamil Internet Conference*, October 2009: 239-243.
3. Nagao, M., Tsujii, J. and Nakamura, J. 1985. "The Japanese Government Project of Machine Translation," *Computational Linguistics* 11 (2-3): 91-110.
4. Nirenberg, S. (Ed.) 1987. *Machine Translation - Theoretical and Methodological Issues*, Cambridge: Cambridge University Press.
5. Rambow, O. and Satta S. 1996. "Synchronous Models of language," *Proceedings of the 34th Conference of ACL*, June 1996: 116-123.
6. Sato S. and Nagao, M. 1990. "Toward Memory-based Translation," *Proceedings of the 13th Conference of COLING*, August 1990, Vol.3: 247-252
7. Shieber, S.M. and Schabes Y. 1990. "Synchronous Tree-Adjoining grammars," *Proceedings of the 13th Conference of COLING*, August 1990: 253-258.
8. Sumaja Sasidharan, Loganathan R, and Soman K P. 2009. "English to Malayalam Transliteration Using Sequence Labeling Approach", *International Journal of Recent Trends in Engineering*, May 2009: Vol 1, No.2
9. Takeda, K. 1996. "Pattern-Based Context-Free Grammars for Machine Translation," *Proceedings of the 34th Conference of ACL*, June 1996: 144-151.



5

கணினியில் தமிழ்ப் பேச்சு மற்றும்
சொற்பகுப்பு ஆய்வு



Bilingual TTS for Tamil and English

AG Ramakrishnan, Vikram LR, Abhinava, ShivaKumar HR

Medical Intelligence and Language Engineering (MILE) Laboratory,
Department of Electrical Engineering, Indian Institute of Science, Bangalore 560012.
agrkrish, vikram.ckm, abhinav.zozo, shivahr@gmail.com

Abstract

An unlimited vocabulary text-to-speech engine has been developed, which currently handles both Tamil and Kannada Unicode text. The input text is processed by a grapheme to phoneme converter module, which uses language specific pronunciation rules to convert the text into an unambiguous phonetic representation. This text is then parsed into demisyllable like basic units. The occurrence of these basic units are searched for, from the phonetically rich spoken database, which is segmented and annotated at the phone level. An unit selection algorithm then selects the best combination of the available speech units to be concatenated to synthesize the speech, which is then converted into .wav format.

Introduction

Text to Speech (TTS) synthesis is an automated encoding process, which converts the given text in a specific language into speech. Till date, only English and some European language TTS systems have gained commercial importance due to their quality output. This paper primarily deals with developing a modular, unit selection based TTS framework for Indian languages. Bilingual Tamil and English TTS is developed for this purpose. However, this framework can be easily modified for any other language. The TTS framework developed is concatenation based, with polyphone taken as the unit of concatenation. This framework is further optimized to suit embedded applications like mobiles and PDAs.

We designed and developed corpus-based concatenative Tamil speech synthesizer in Matlab and C. A concatenation based speech synthesizer requires a rich and large speech database with varied and natural distribution of prosodic and spectral characteristics of speech sounds. The sentences to be recorded need to be selected from a text corpus. We used CIIL (Central Institute of Indian Languages, Mysore) Tamil text corpus for our research. A greedy algorithm is used to select phonetically rich sentences from this huge corpus. This resulted in 1026 sentences, which were recorded from a professional, native Tamil speaker. These sentences are segmented offline and the database is organized in such a way that it facilitates faster search.

During synthesis, from the phonetic transcription of the sentence to be synthesized, specifications of the required target units are predicted. Units are then selected from the database that best match the target specification according to a distance metric and a concatenation quality metric. These units are then concatenated to produce synthetic speech. There may be audible glitches in the output after concatenation. This could be because of either poor segmentation of the speech database or improper selection of units by the TTS frame work. In our case, we know that the segmentation is nearly error-free. Hence, post-processing is performed on the final set of units. This includes smoothing the pitch contour,

during concatenation, at junctions of units with unacceptable pitch discontinuity. Our experiments reveal that about 15-20% of the unit junctions require pitch smoothing. Optimal coupling technique is then used to concatenate these units at appropriate positions. This resulted in intelligible and reasonably natural synthetic speech.

Intelligibility of synthetic speech also depends on selecting the units that match the target phonetic contexts. At times, the required phonetic context may not be available in the database. In such case, we propose that similar phones that are perceptually indistinguishable may replace these phonetic contexts. The most confused pairs of Tamil phones, which can be replaced by each other in specific contexts at the time of synthesis (if they are not available in the corpus) are found. Explorative experiments to determine the applicability of incorporating these techniques resulted in high mean opinion scores for the synthesized output from the native Tamil evaluators. Hence, we consider that this possibility of replacing missing phonetic contexts can be used in practical TTS.

Finally, when any person speaks the same sentence repeatedly, the speech waveforms don't have identical characteristics. With this motivation, the final portion of my research attempts to analyze the variability of characteristics of different instances of speech, when a speaker utters the same sentence multiple times, at different times. The idea is to look at the possibility of generating a slightly different synthetic speech each time the same text is synthesized, thus trying to make the TTS sound not monotonous and more human like.

Also, we observe that incorporating prosody and pause models for Indian language TTS would further enhance the synthetic speech quality output. These are some of the potential, unexplored areas ahead, for Indian speech synthesis.

Motivation for bilingual TTS

In the present scenario, usage of English in Tamil text has become common and inevitable. If such words are omitted in TTS, the TTS would be less effective. Hence we have developed a bilingual Tamil TTS for generating Tamil and English by using the same synthesis Tamil data for both the languages. The sparsely occurring English text is converted into phonemes using a separate grapheme to phoneme converter and the corresponding phonemes are obtained from the available Tamil database for concatenation. The initial results are encouraging and we are working on some more improvements for better sounding English.

Tamil synthesis database has 5 hours of Tamil sentences recorded by a male professional Tamil speaker. The recorded database is rich in phonetic context and phonetic variations. The TTS takes Tamil Unicode input, and performs equivalent phonetic translation. In Tamil, some letters and phonetic contexts influence the phoneme of a letter, such as "ka" can become "ga" in some phonetic contexts. The rules for these phonetic changes are coded as rules in the grapheme to phoneme converter. The phonemes are then again grouped into polyphonic cluster. The clusters are searched in database to find a best choice is selected for a given context. We have developed a prosody based unit selection algorithm to further enhance the best selection for a given text.

Features/Specifications of the TTS

1. Unlimited Vocabulary : Any sentence involving any combination of native words of Tamil (or Kannada) is handled.

2. Quality : The intelligibility of the TTS is quite high, and it is also acceptably natural.
3. Text Encoding: Only text entered in Unicode will be handled. Other proprietary or public font encodings are not handled, and will not be handled, even in the future.
4. Web Demo: The TTS has been made available as a web demo. Go to the link, http://mile.ee.iisc.ernet.in:8080/tts_demo. The TTS demo page can be seen and a box, where the Tamil or Kannada text in Unicode must be submitted.
5. Test Input to the TTS: If you want to submit your own custom text input, you can do so, by typing using our open source Multilingual Indic keyboard interface: பன்மொழி வாயில் or Vishwavaangmukha), which you can download from <http://code.google.com/p/indic-keyboards>
6. Output Format: The TTS outputs a standard .wav file.
7. Testing: The web demo of Tamil TTS has now been tested by hundreds of people from around the world, and many GB's of synthesized .wav files have been downloaded by them.

Current Limitations

It cannot handle numerals, proper nouns and words originating outside the current language handled by the TTS, sentences needing intonations changes, such as interrogative and exclamatory sentences. There may also be some mispronunciations at times. In the case of long sentences, even the pauses (phrase breaks) may not be at the instances, where a human speaker will naturally pause for better intelligibility and clarity.

Future Enhancement

All the current limitations mentioned above are being addressed by us, and will be updated in time. The numeral handling facility will be added very soon. Some of the causes for mispronunciations have been identified, and will be corrected next. However, further enhancement of naturalness requires significant research in computational linguistics and prosody and hence, may take considerable time to reach good quality.

Conclusion

MILE Laboratory has teamed up with Bookshare.org, an International non-profit organization, to provide Tamil and Kannada digital books (copyright free or permitted by authors) online to print-disabled people (visually challenged, old people with vision disabilities and people with other disabilities that make it impossible for holding a book and turn pages of it). The MILE OCRs and TTS in the respective language (Thirukkural / Vak) will be used for this purpose, and thus the printed content can directly be heard as speech on a desktop computer or laptop.

Acknowledgment

The TTSs were developed as part of a Research Project, funded by Ministry of Social Justice and Empowerment (MSJE), Government of India. The author also thanks Karthigeyan, Jayavardhana Rama, Prathibha, Muralishankar, Lakshmith Koushik, Parthasarathy, Arun, Abhinav, Vikram, Shivakumar and others, who contributed to the TTS research and development over a period of time.

Voice based Tamizh Tutoring System for Primary School Children to Enhance their Tamizh Learning Interest

P. Vasuki and C. Aravindan

SSN College of Engineering

Chennai, TamilNadu, India

VasukiP@ssn.edu.in, 2AravindanC@ssn.edu.in

Abstract

In this paper, we focus on interactive tamizh learning system for children. Our system displays a picture and asks the child to recognize the picture and to speak out its name. Our speech engine captures and analyses the speech to give proper feedback. The speech engine has been trained with a corpus of words to obtain probability distributions (templates) of spectral coefficients of words. Spectral coefficients of the input speech are extracted and the word is recognized by matching the spectral coefficients with the templates. This word is then compared to the original text of the picture being displayed. By making many trials at different levels, the system induces the child to utter the object identity in tamizh. A scoring system runs at the background to calculate scores based on number of fouls and success. Providing picture along with the word helps the child to easily register the word in written form and meaning of it. The child who receives such training will acquire quick tamizh reading and speaking capacity. The conceptual mapping of images with words improves the declarative knowledge of a child.

Keywords – Speech Recognition, Computer Based Training, Tamizh Tutoring System, and Hidden Markov Model

1 Introduction

Pure class room teaching may be monotonous and computer based training acts as good supplement to it. Computer based training becomes more effective when it has a speech interface. Such systems are popular in foreign countries, but not available in tamizh, as there is less focus in tamizh speech processing. The system that we have designed provides a friendly environment for children and enables them to talk to the system in tamizh, while learning tamizh. Thus our system may help to motivate learning of tamizh & makes it joyful and interesting.

The Declarative knowledge is factual or experimental [7]. Displaying an object, pronouncing it, displaying its text and make a child to realize and later recognize improves declarative knowledge of the child and enables the child to map the word with real meaning of it. The child is active while learning by using her psychomotor skills, i.e. the child talks to the system using mike, uses keyboard and mouse for navigation and also using her cognitive skills to recognize the picture by looking it, recall, utter it correctly, learn to utter, learn to read.

The system gets speech input using omni directional mike and records the speech in wav format. The speech characteristics vary at each instant, each speaker even for same word. Thus we have gone for extensive training with ten speakers of various age groups and ten utterance of each word; i.e. every word is trained with hundred instances of utterance. Likewise the system is trained with hundred and twenty five basic tamizh words which are collected from first standard tamizh book.

HTK toolkit (Hidden Markov Model Toolkit) is used for building the speech engine[6]. The speech samples are parameterized in the form of Mel Frequency Cepstral Coefficients (MFCC) which reflects the characteristics of perception. In training, the word is divided into 'n' states. For each state an acoustic model is generated based on Gaussian Mixture Model (GMM). The probability of each state initialized and re-estimated based on its sequence using Forward Backward algorithm. HMM Model is generated for each word as initial state probability and sequence of states with transition probability.

In testing, MFCC features of testing word are generated and mapped against the model of each and every word in the vocabulary set, using Viterbi Decoding Techniques. The training wave gets the label of the word that closely matches with highest score. If the word uttered is not the original name of picture, it plays a sound of it with the help of look up table. A scoring system is running at the back to calculate the score and when the child quits, the score is displayed.

Section II talks about the system architecture in brief and section 2.1 briefs about speech Engine , 2.3 describes about tamizh tutoring system and section 3 briefs about scope and future enhancement.

2 Architecture

The project is of two phases (i) tamizh Speech engine (ii) Tamizh Tutoring system.

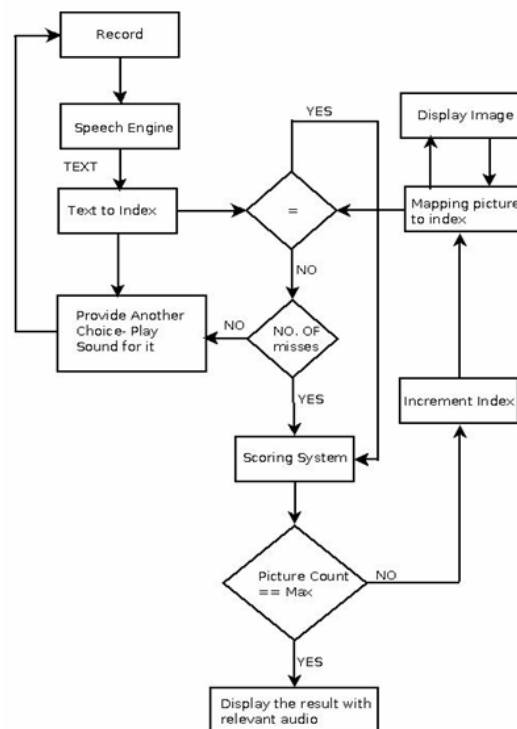


Fig.1. Architecture

In first phase, Speech Engine has been developed with extended training with different speakers and various words. The speech Engine recognizes the word and converts into text. In second phase is tamizh tutoring system is implemented with speech interface on the top of speech engine. Fig. 1 shows the architecture of the complete system.

2.1 Speech Engine

Our Speech engine is an Isolated Speech Recognizer (ISR) implemented using Hidden Markov model. It takes isolated word as a processing unit recognize it and produces text. The system's word error rate is less when tested against system gets input as isolated word than connected word recognition. Our tamizh tutoring system needs only isolated speech recognition.

A. Data Preparation

i) Corpus Collection

The speech Corpus has been collected from 10 speakers of different ages(6 year to 44 years) and different gender with sampling frequency of 16 khz and at 16 bit quantisation rate. They read 120 words in neutral form. Every speaker utters each word 10 times. The words are annotated manually. The signal is pre-processed before training.

Windowing is done to avoid the discontinuities of frames at starting and end [2]. The signal is divided into frames and windowed using hamming window of size 20 ms with the overlapping window size of 15 ms.

$$S_n' = 0.54 - 0.46 \cos(2\pi(n-1)/N-1)S_n \quad \text{-----(1)}$$

ii) Feature Extraction

For each frame MFCC Mel frequency cepstral Coefficients are calculated. MFCC is a representation of the short-term power spectrum of a sound, based on linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency [4].Fig. 2. shows the processing steps of deriving MFCC from speech.

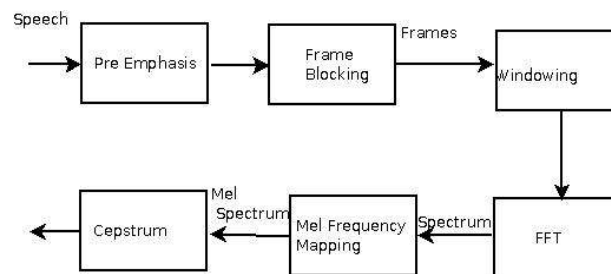


Fig 2 :Derivation of MFCC From Speech samples

MFCC –Mel frequency coefficients are working based on perception of word by human ear. Mel frequency Cepstral coefficients are 12 filters output that have been collected from the short time frame.

MFCC Feature vector of size 39 represents the signal of the frame.13 parameter value (12 MFCC coefficients and absolute energy of the frame), 13 delta coefficients (variation of cepstral coefficients

between consecutive frames) and 13 acceleration coefficients are calculated for each frame. Fig 3 shows the screen shot of MFCC collected from different frames of a speech sample.

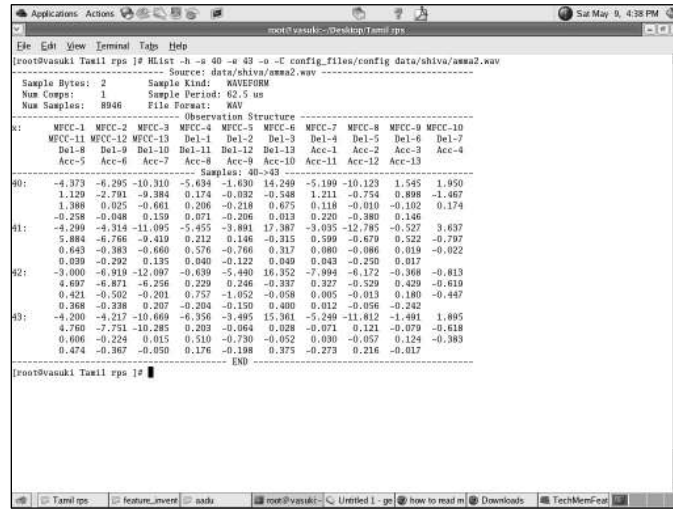


Fig 3. Sample MFCC vectors derived from few frames of a speech sample.

B. Word Modelling

The quantized MFCC parameters are used to construct a HMM model for each word.

HMM – Hidden Markov Model is as an efficient approach for short time stationary stochastic sequence [4]. Thus speech signal is modelled using Hidden Markov model[9].

In HMM training, HMM Model is created for each and every word in the training set. The Model is $\lambda = (\Pi, A, B)$, where Π is the initial probability distribution, A is state transition probability and B is output probability. For a class whose data is considered to have multiple clusters, a Gaussian mixture model can be used to represent its probability distribution. The signal is represented by $N(\mu, \sigma)$. The output probability is estimated using Gaussian Parameter (μ, σ) . The state transition probability is calculated based on state changes.

C. HMM Parameter Initialization

The word should be divided into number of sequential states. Number of states is fixed for each word. We use three utterances of a word to form a mixture. HMM parameter template is created with initial values of mean 0.0 and variance is 1.0 for the 39 feature vectors.

Gaussian Mixture model is used to find the probability distribution of the state of every mixture using expectation Maximisation algorithm. The signal is represented using Gaussian distribution using the formula [4]

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{-----(2)}$$

D.HMM Parameter Re Estimation

The word is divided equally into number of states based on timing. The frame is expected to belong to a certain state. With that expectation the mean and variance of every state is calculated. The probability of each frame belonging to the corresponding state specified is calculated based on the mean, variance of that state and feature value of that frame. If the probability maximize the frame belongingness in that specified state the frame is retained there else moved to the other state where its probability of belongingness is more. Thus the HRest tool used to re estimate HMM parameters using Baum Welch Algorithm. Baum Welch algorithm estimates the probability in the forward backward sequence. i.e. the probability of n'th frame is calculated from the previous n-1 frame, in Backward algorithm the probability of n'th frame is calculated from the remaining n+1 to N frames. HMM Parameters are defined based on forward and backward probabilities. An acoustic model has been generated for every word as a sequence of states.

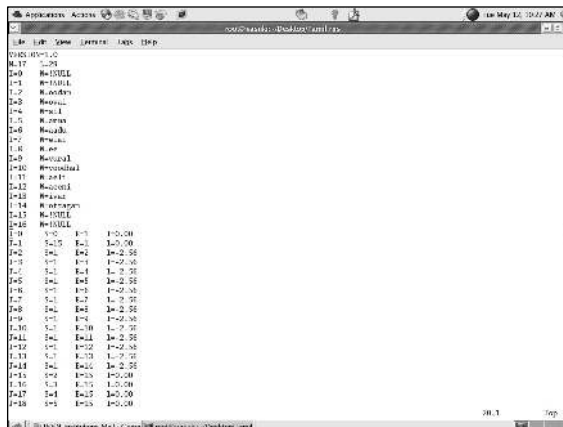


Fig. 4. Re estimated values HMM Parameter after applying forward backward algorithm.

E. Word Net (Language Model) formation

Word net describes about the different sequence about the word occurrence and the probability of the sequence [5]. The apriori knowledge about the sequence of states improves the performance of speech recognition system [10]. Dictionary defines the grammar about the sequence of word occurrence. Our system recognize only isolated words, thus dictionary is formed such as individual word as single entity.

Fig 5 Word net – Network of words formed between 12 tamizh words.



The word net is represented in SLF, ie Standard Lattice Format in terms of nodes and arcs. Node represents word and arch represents link between words. In the current word net we have 120 words (including NULL words) and arcs among the word. Fig 5 shows screen shot of sample file in Standard Lattice Format developed with 12 words network.

A statistical word net is formed by having words are as nodes and the link -log transition probability as the link between arc.

F. Recognition

HMM recognizer recognizes the words based on the language model (SLF) and acoustic model created by training set. Initially forced alignment is used to align the state of the testing utterance [5]. The Viterbi decoding algorithm matches the state with the trained states. At every step, P suitable states are selected among the different states of M models. Then it produces the sequence of the next many possible states. Among those many possible states few highest probable states are chosen and so on. Likewise the lattice structure preceded and compared with the training pattern. For every possible state in every model the probability of observation is calculated. The overall probability of all possible states per model is multiplied. The model with the highest overall probability will be chosen as the resultant model (word) of utterance.

The word uttered is recognized based on Viterbi decoding techniques[8]. The HMM parameter of the state sequence of uttered word is generated and matched against the sequence of trained words as shown in the figure 6. Many similar words have the higher probability and the word is labeled as label of word which has the nearest matching or has highest probability [5].

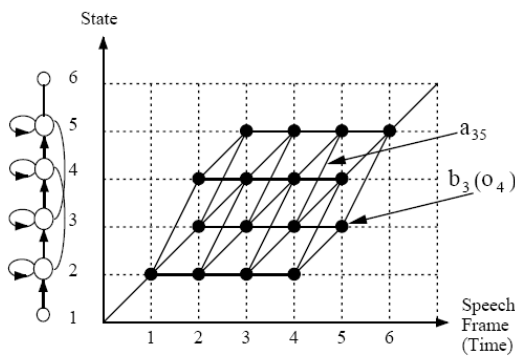


Fig : 6. Vitebi Decoding Alogithm L.R.Rabiner and B.H.Juang, Fundamentals of Speech Recognition, Pearson Education, 1993.

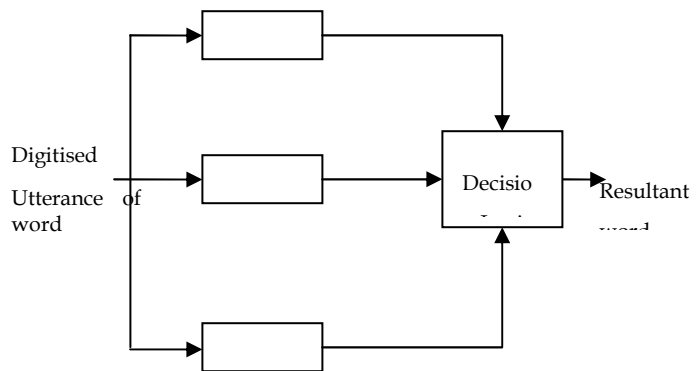


Fig 7. HMM Model for Isolated word Recognizer – The observation vector applied to HMM models of different words.

The probability of input belonging to different models are compared in Decision Logic. The model with highest score is the model of input observation.

2.2 Tamizh Tutoring system

The tamizh tutoring system is devised to enhance the tamizh learning interest amongst children. A Picture driven system along with audio attracts the children to learn tamizh language. It bridges the gap of lack of interest in learning with usual keyboard and mouse interface. Therefore there will definite impact in education system on this interactive tutoring.

This tutoring system has been developed with inputs from various theories of learning such as Bloom's Taxonomy [6]. According to it, the learning consists of three domains. Cognitive, Affective, and Pshycomotor skills. Cognitive learning is improving the mental ability (Informative, Recognizing and Recalling) - our system brings out cognitive skills by making the child to recognize an object, recall it. It addresses Affective skills (receiving, responding valuing, motivation and appreciation) by responding on her recognition. It also brings in Psychomotor skills (involve more physical activities) - by making the child to use microphone, mouse keyboard and speaker.

A picture of index one is displayed and a child is asked to name the picture. The child has to record the identity of object in his voice. The word is recorded and given as the input to speech engine. The speech engine produces the corresponding word in textual form. The word is mapped into index and matched against the index of the picture displayed. The matching module sends a signal to scoring system. The scoring system evaluates and pop out choice for next trial if needed The system gives maximum three choices per word to a child, if the child recognition is wrong or uttered the word with different pronunciation, the word is displayed to child in textual form and audio corresponds to the word is playing. On success, the system just display the object identity in text then system iterates through next picture till end. At the end of picture gallery the system displays score with the word of appreciation and continue till the child wish to quit. This system not only trains the oral pronunciation of word but also trains the object in written form. Figure 8 shows the screen shot of VBaTTS system.



Fig:8 VBaTTS -Screen Shot – Ready to record a child voice on “padivu sai” button and “aduthu” button helps to navigate on picture gallery.

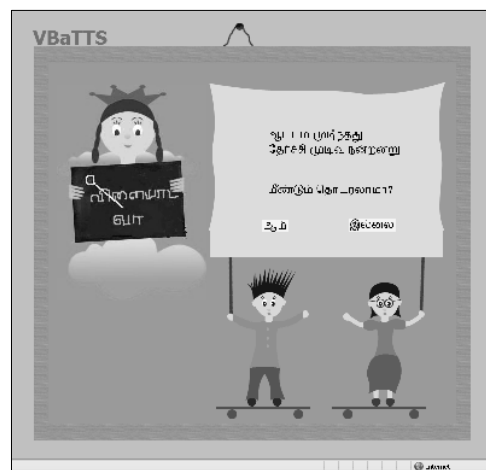


Fig 9.VBaTTS Evaluation sheet. It shows screen shot Final evaluation sheet which displays the evaluation result is not good and asks child wish to continue or not.

The scoring system adds score based on number of foul and negates score on failure. Child's satisfaction or motivation or feeling of competence is also important. The child's achievement level is meaningless if the child chooses not to exercise the skill after acquiring it. Thus the system operates in different level and the score is displayed in grade if it is Excellent or very good else the system just says the score is not good enough and asks the child to reattempt.

Conclusions and scope for future work

The speech engine is not producing 100% accuracy. To overcome this, system has to be enriched with speaker normalization and emotion normalization. The accuracy of the system also can be improved by adding some more parameter like speech rate and intensity etc to build the acoustic model of a word.

Acknowledgment

We acknowledge Dr. T. Nagarajan, Professor and Head, Department of IT, SSN College of Engineering, Chennai, for his valuable guidance in developing speech engine. We also thank AICTE for funding a project on development of tamizh speech engine.

References

1. Ben Gold and Nelson Morgan, *Speech and Audio Signal Processing*, John Wiley and Sons Inc. , Singapore, 2004
2. L.R.Rabiner and R.W.Schaffer *Digital Processing of Speech signals* Prentice Hall -1978
3. Veenayagamoorthy, Viresh Kumar and Kumes Sadraseragan *Voice recognition using neural network in Science Direct - IEEE* 1998
4. L.R.Rabiner and B.H.Juang, *Fundamentals speech Recognition*, Pearson Education, 1993.
5. L.Rabiner *A Tutorial on Hidden Markov models and Selected Applications* IEEE, Vol. 77 No. 2, 1989.
6. Benjamin S. Bloom *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain ISBN-13: 9780582280106* Publisher: Addison Wesley Publishing Comp
7. M.Helander, T. K. Landauer, P. Prabhu (Eds) *Intelligent Tutoring Systems -Handbook of Human-Computer Interaction*, Second, Completely Revised Edition, Elsevier Science B. V., ©1997, Chapter 37
8. Schultz, W.Black, Schultz,W.Black,Horynyak,Kominek, *SPICE :Web Paged Tool for Rapid Language adaptation in speech processing systems*, in the proceedings of INTERSPEECH 2007
9. Krstulovic, S., Hunecke, A. and Schrder, M. (2007). An HMM-Based Speech Synthesis System applied to German and its Adaptation to a Limited Set of Expressive Football Announcements.Proc. Interspeech Antwerp, Belgium
10. Saraswathy, S., Geetha T.V. 2004. Building language models for Tamil speech recognition system. In Proceedings of AACC2004 and LNCS3285, 161-168

Prediction of Pauses in TTS - Tamil

P.Arulmozhi

Project Associate, MILE Lab
Department of EE, IISc, Bangalore.
arulmozhi@ee.iisc.ernet.in

A G Ramakrishnan

Professor, Department of EE
IISc, Bangalore.
ramkiag@ee.iisc.ernet.in

Abstract

Text to Speech (TTS) involves the task of converting text typed in electronic format into a speech signal. In MILE lab, we are involved in making a TTS system for Tamil and Kannada. In this paper, the contribution of syntactic information such as Part of Speech (POS) Tags in enhancing the quality of a Text to Speech (TTS) synthesis system for Tamil is researched. The quality of a TTS system is measured by the intelligibility and naturalness of the synthesized speech. The NLP module of the TTS system (for example, text normalization) contributes to not only to its intelligibility, but also to its naturalness, by improving the prosody. The stress and pause modeling can be improved using the POS and other syntactic information. In a sentence, where there should be a pause, and where there should not be a pause has to be identified for the naturalness of the produced speech. Because, a sentence without pause or with identical pause intervals sounds robotic. Also, pause at a wrong place makes the sentence unnatural and there is a possibility of change of meaning. For example take the following sentence,

avarukku inRu <P> mAlai kitaittatu.

avarukku inRu mAlai <P> kitaittatu.

<P> here indicates that there should be a pause. The pause given in different places give different meanings. Syntactic information such as Parts of Speech (POS) can be used for identifying the rules for pause in a particular sentence. A rule based POS tagger is developed for this purpose without using a root word dictionary. Currently, manual evaluation shows an accuracy of approximately 74% using only the lexical rules. The performance is expected to improve after the context sensitive rules are applied. Rules are made for predicting the insertion of pause at the right place. The manual evaluation of pause insertion shows a significant improvement in the naturalness of the produced sentence.

1. Introduction

This paper presents a rule based Parts of Speech (POS) tagging method in the perspective of improving the naturalness in the synthesized speech. The quality of a TTS system is measured by the intelligibility and naturalness of the synthesized speech. There are two main modules in a TTS system. One is the Natural Language Processing (NLP) module, which takes care of the production of phonetic transcription, intonation and rhythm. Another is the Digital Signal Processing (DSP) module which takes care of the production of the speech wave-form of the given text. The stress and pause in the speech contribute majorly to the naturalness, which is controlled by the NLP module. Using this POS tagger, we try to find out the right place to introduce pause in the synthesized speech. Introducing a high degree of naturalness is theoretically possible, but the rules to do so are still to be discovered (Jonathan A. 1996

Many linguistic aspects are analyzed (Thierry D) and syntactic information such as POS tagging is considered important to achieve a good TTS. In this paper, we present a POS tagger created for Tamil, which is a highly agglutinative and partially free word order language.

2. POS Taggers

The purpose of a POS tagger is to automatically find out the syntactic category of a word in a sentence. Different methods may be followed to do POS tagging. Most commonly used are rule based, statistical, and transformation based methods. Rule based taggers work on predefined linguistic rules for deciding the syntactic category of a word in a sentence. These rules may be lexical or context sensitive and they are language dependent. Statistical taggers work on the calculated probability, in which a POS tag for a particular word is decided based on the lexical and contextual probability. A training corpus is used to train the system and the input sentence is tagged based on the probability which is calculated using the training corpus. Transformation based taggers derive rules based on learning. Those rules are used to find the syntactic category. In English, Brill's tagger is the most commonly used, TBL based tagger. There are statistical, rule based and hybrid taggers being worked on for Tamil (Arulmozhi P, Sobha L, 2006).

3. MILE POS Tagger for Tamil

3.1 Purpose of MILE tagger

In natural speech, we introduce stress and pause at the right places, so that it is easily understandable. In a TTS system, the DSP component produces the speech wave-form and the NLP component is responsible for the naturalness of the produced speech. It should identify the right places to introduce pause. The amount of pause to be introduced has been identified. The pause is introduced using the category of the words in a sentence. Other syntactic information such as shallow parsing and clause boundary identification will be helpful not only in identifying pause, but also to find out the pitch and intonation.

There are a few POS taggers available for Tamil. They have been developed for the purpose of preprocessing for NLP activities such as machine translation and information extraction (Arulmozhi P. et.al 2004). The probability based POS taggers need huge training corpus. They provide the POS tags according to the tagset used for training. In the case of a TTS, we do not need such detailed tagging.

3.2. Nature of Tamil Language

Tamil is a morphologically rich, agglutinative and partially free word order language. Compound words are common in this language, where two or more words are combined to form a single word. The case markers and tense markers appear as inflections of the root word itself. For example, taking the word 'varukirAn', the inflections and root word can be split as follows.

vA + kir + An

vA - root word

kir - present tense marker

An - 3rd person, Singular, Neuter gender.

Tamil is a partially free word order language because changing the word order to some extent does not affect the meaning of the sentence. However, this order change can-not occur within a phrase. For the sentence,

1 'Aciriyar nanRAka paTitta mANavanukku paricae kotuttAr'

Teacher thoroughly studied Student+Dat prize+Acc gave+Hon

The teacher gave the prize to the student who studied thoroughly.

It can be written as

1 'nanRAka paTitta mANavanukku Aciriyar paricae kotuttAr'

2 'Aciriyar paricae nanRAka paTitta mANavanukku kotuttAr'

but,

1 * 'Aciriyar paTitta mANavanukku paricae nanRAka kotuttAr'

changes the meaning. So, within a phrase, the word order does not change.

3.3. Tagset

A tagset is the set of all tags used by the POS tagger. There are two levels of tags - the main tags and sub tags. The main tags identify the main category of the word such as noun verb or adjective. The subtags identify the category of the inflections such as person, number, gender, and tense. For the purpose of TTS, we do not need very detailed tags unlike other NLP activities. But using only the main tags would not give sufficient information. So we need some of the sub tags too. So, as a special case, we have developed a tagset for the purpose of inserting pause in a sentence. In our tagset, each tag is a combination of a main tag and one or more sub-tags. The nouns take the case and plural markers and the verbs take person, number, gender, and tense markings. Apart from this, pronouns have person, number, and gender. The clitics are suffixed to the root word to form adverbs and conjunctions. Then the dates, numbers, and punctuations are also tagged separately. English POS taggers and some of the Tamil taggers (Dhanalakshmi V. et. al. 2009) use monadic tags. Monadic tags do not give information on inflections, which is important for TTS. We use structured tags such as "NN+pl.acc" in which different pieces of information serve in different parts of the rules. This tag means, a noun in with the inflection for plural and accusative. We have 15 main tags and 30 subtags adding up to 45 tags.

3.4 MILE Tagger:

POS Tagger. This POS tagger is a rule based one. We do not use a root word dictionary. The tagger is based on a two stage architecture. The block diagram of the POS tagger is shown in Figure 1.

In the block diagram, each block explains its functionality. The first stage has the lexical rules and the second stage has the context sensitive rules. First, a sentence is taken as input and split into tokens. For each token, the suffixes are identified. Then, using the lexical rules, which work at the word level, each word is assigned a POS tag according to the suffixes identified. Then, this output is given as an input to the second stage, where the context sensitive rules change the tag if it is wrongly tagged by the lexical rules. Thus, the final tagged sentence is obtained.

Separate tables are created for programming purpose with the list of suffixes identified. A lexical rule looks like,

2*1+1*1, NN+pl.acc

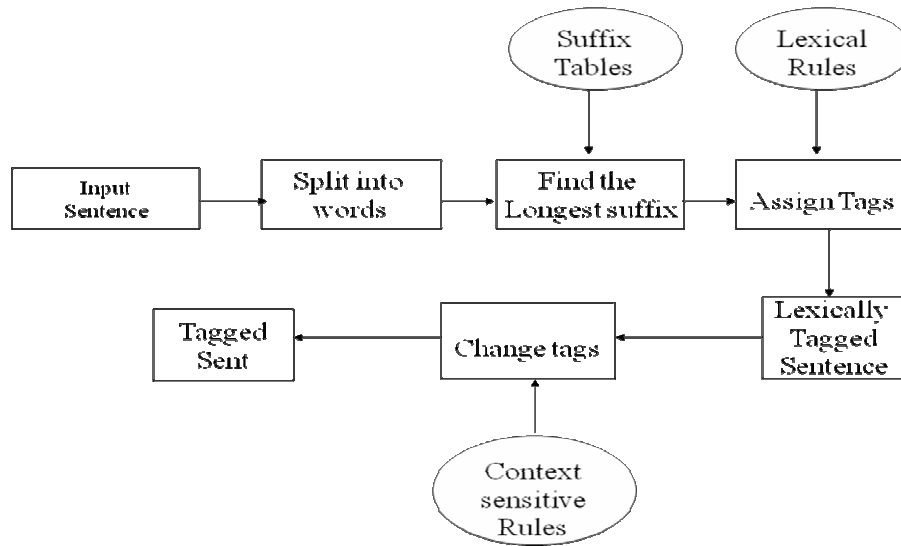


Figure 1

This means, the suffixes indexed 2*1 (Suffix Table 2 Column 1 - kaL) and 1*1 (Suffix Table 1 Column 1 - ae) occur in a sequence, the word will be tagged as Noun+Plural+Accusative. Here 'kaL' is the plural marker and 'ae' is the accusative marker. There are 13 such tables which list 103 suffixes identified and put in. These suffixes are used by the lexical rules. The context sensitive rules are embedded in the system. For example the following can be considered as a context sensitive rule.

‘If a sentence starts with a verb, change it to noun’.

Since Tamil is a verb ending language, a sentence can-not start with a verb. So, if the first word of a sentence is wrongly tagged as a verb in the first level, it will be corrected in the second level. The combinations of lexical rules including the inflections are 533 and the number of context sensitive rules are 4.

For any POS tagger to work correctly, the sentence boundaries have to be identified. We use a sentence splitter for splitting paragraphs to sentences. Input to the sentence splitter is any Tamil text such as paragraphs. The output is an array of sentences. This process is also embedded in the POS tagger based on our need. We use a rule based sentence splitter and the rules are heuristic in nature.

4. Pause Model

Pause model. Insertion of the right amount of pauses at the right places adds to the naturalness of the synthesized speech. With a natural language text, native speakers introduce pauses with the knowledge of the language acquired. But in a TTS system, those pauses need to be inserted by the system at the right

places. For European languages such as Spanish, there are rule based pause models developed and experimented (Rafael M 2002). A wrong pause inserted between two words may make the synthesized speech unnatural. For simplicity, such an example sentence is illustrated in English. Here, the notation <np> denotes “no pause” between the words, whereas <p> refers to the required pause.

Example:

The< np> book<np>is<p> on<np>the<np>table.

The<np>book<np>is<np> on<p>the<np>table.

Speech synthesized as per the tags for the first sentence will appear perceptually natural, whereas for the second sentence, a wrong pause between words ‘on’ and ‘the’ will appear perceptually unnatural. Hence POS information and pause are very important in the context of TTS. In this paper, we focus on pause insertion between successive words by pause prediction from the POS tags estimated from the input text. Presently we use key words and heuristic rules for inserting pauses. Rules for inserting pauses at the right places are created according to the POS tags.

At present, in the syntactic level, only POS tags are considered for identifying pauses. It is considered as a basic preprocessing needed. On top of this, the phrase chunks may be identified, and that will be useful in identifying the places where pause must be inserted. In that process, the phonological phrases must be identified finally for identifying pauses and intonation.

Six levels of pause have been identified, which determine the duration of the pause. In Chinese pause model, they use minor, major, and punctuation breaks (Fu-chiang et.al 1997). We have defined <P0>, <P1>, <P2>, <P3>, <P4>, and <PW>. P0: No pause, P1: Lowest pause, P2: Medium Pause (Ex: Pause after a comma), P3: Significant Pause (Ex: Pause after a semicolon), P4: Highest Pause (Ex. Pause between sentences). The pause levels P0 to P4 are derived from the existing synthesis database. This work has been carried out for Tamil and can be extended to all Dravidian languages with no major change. <PW> is the common pause between each word. Wherever <P0>... <P4> is not identified, <PW> is assumed.

At the initial level, <PW> is assumed for each word, and rules for converting it to <P0> to <P4> is made. The sample rules are given below.

1. There is no pause (or may be very minimal pause) between a number and certain words following, such as ‘mani’, ‘latcam’, ‘kOTi’. There is a list of words defined for this rule. Any noun in plural form, (NN+PL) after a number does not have a pause.
2. If the previous word has an accusative/dative marker, and the current word is a postposition, there is no pause between the current and the previous words.

Ex : avanai <P0> pola,

avanukku <P0> pin

3. Combine the words with POS tags Adjectival Participle (AJP) and Noun (NN - any number of them) occurring together. There is no pause between them.
4. There should be a pause before quantifier (Q).

Ex : (azhakiya kiraamamum)<P3>(oru periya ooraaTciyum)

There are 15 such rules made for converting <PW> to <P0> - <P4>. In natural speech, we do not give any pause between most of the words. Taking that into consideration, we did another experiment. Instead of taking <PW> as default pause, we took <P0> as default ie. There is no pause between any word initially, and then inserting pauses using rules. This reduced the rule set from 15 to 8, because we had more rules for <P0>. The output from both is obtained and given for evaluation.

5. Output and Evaluation

The output of the system contains the original sentence, the predicted POS tags and the predicted pause levels. All are displayed parallelly. An example output is shown in Figure 2.

சி.பி.எஸ்.இ.	முறையில்	படிக்கும்	மாணவர்களுக்கு	அடுத்த		
C.B.S.E	stream+loc	studying	students+pl+dat	next		
NN	NN+loc	VB+fut+3sn	NN+pl+dat	AJP		
<P0>	<PW>	<PW>	<PW>	<P0>		
கல்வி	ஆண்டு முதல்	10ம்	வகுப்பு தேர்வு	கிடையாது		
education	year	from	10th class	exam No		
NN	NN	PP	Q	NN	VB+pst+3sn+neg	
<P0>	<P0>	<P2>	<P0>	<P0>	<PW>	<P4>

Figure 2. Example Tamil Sentence and Output

In the figure, the first line is the Tamil input; second line is the corresponding meaning in English; third line is the predicted POS tags and the fourth line is the pause levels identified.

This output is given to the DSP module and the wave form of the sentence obtained. The people who evaluated the outputs are native Tamil speakers, who did not have knowledge about the methods used for creating TTS outputs. 3 types of outputs are given to the evaluators and Mean Opinion Score (MOS) is obtained. 10 sentences are spoken by the TTS as follows:

Without implementing the pause model

After implementing the pause model with default as <PW>

After implementing the pause model with default as <P0>

The scores are given from 1-5 according to the understandability and naturalness (1-worst, 5-best). The evaluation based on the mean opinion score gives encouraging results.

The rule based POS tagger is evaluated manually for the correctness of the tags. The evaluation is done manually. In a given tag, if the main tag is correct and the sub-tag is wrong, or vice versa, we take that as a wrong tag. The system gives 78% results without a root word dictionary. More context sensitive rules are to be added so that the accuracy of the POS tagger also can be improved.

References

1. Arulmozhi. P, Sobha. L, Kumara Shanmugam. B. 2004. Parts of Speech Tagger for Tamil. In the proceedings of Symposium on Indian Morphology, Phonology & Language Engineering, (March 19-21) Indian Institute of Technology, Kharagpur (Page 55-57).
2. Arulmozhi Palanisamy and Sobha Lalitha Devi. 2006. HMM based POS Tagger for a Relatively Free Word Order Language. A poster presentation in CICLing-06 (February 19-25) at Mexico.
3. Brill, E. 1995 Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging. *Computational Linguistics*, 21, 4 (Page 543-565).
4. Dhanalakshmi V, Anand kumar M and Soman K P, 2009. POS Tagger and Chunker for Tamil Language. Tamil Internet Conference, Cologne, Germany. pp. 250-255.
5. Fu-chiang Chou', Chiu-yu Tseng, Keh-jiann Chen, and Lin-shan Lee 1997. A Chinese Text-to-Speech System Based on Part-of-Speech Analysis, Prosodic Modeling and non-Uniform Units. ICASSP'97, Volume - 2, Munich, Germany.
6. Jonathan Allen 1993. Linguistic aspects of speech synthesis. Presented at a colloquium entitled 'Human-Machine Communication by Voice. Organized by Lawrence R. Rabiner, held by the National Academy of Sciences at The Arnold and Mabel Beckman Center in Irvine, CA.
7. Rafael Marin, Lourdes Aguilar, David Casacuberta, 2002. Placing pauses in read spoken Spanish: A model and an algorithm. *Language Design: Journal of Theoretical and Experimental Linguistics*, pp. 49-67.
8. Thierry Dutoit. High-quality text-to-speech synthesis: an overview. Faculte Polytechnique de Mons, TCTS Lab, 31, bvd Dolez, B-7000 MONS, Belgium.

Text Analysing and Retrieval System using Tamil Phonemes and Vector Space Model

Premalatha.R

rprema79@rediffamil.com

SCSVMV University

Kanchipuram, Tamilnadu, India

Srinivasan.S

sriniss@yahoo.com

Tamilnadu Open University

Chennai, Tamilnadu, India

Abstract

Intelligent information retrieval is one of the important topics in the 21st century. In Tamil documents, generally, morphology (separating noun and verb) concept is used to retrieve the text. In this system we use Tamil phonology so that we can widen our search criteria namely, Vowel – Short, Long; Consonant - Vallinam(Hard), Mellinam(Soft) and Idaiyanam(Medium). So the system would search quickly. Hence the performance of the system would increase. In the classical system, the user should give the exact word to retrieve the information. But in TAR system, as the system internally has its spell checker, the misspelled word can be corrected and also the information can be retrieved. This would be useful for Tamil literates, Tamil students, Tamil learners, etc.,

Introduction

For thousands of years people have realized the importance of archiving and finding information. With the advent of computers, it became possible to store large amount of information and finding useful information from such collections which became a necessity. The field of IR was born in the 1950s out of this necessity. Over the last forty years, the field has matured considerably. Several IR systems [9] are used on an everyday basis by a wide variety of users. The need for information retrieval [1] arises in the Tamil literature documents from the ancient era to the latest, which helps in sharing the data through the internet. In this paper, we are going to do text analyzing and retrieval system using Tamil phonemes and vector space model [8].

In this system, searches can be done in three ways namely (i) Main topics searches (ii) Subtitle searches and (iii) Keyword searches.

The input word would be given in Tamil using Tamil virtual keyboard.

In the classical system, the system [4] read the whole word and it identified whether the word is a noun or a verb. Finally it found the word from the database. In the past it took a lot of time for retrieval. Because in the older system once the entire word is entered, the system would do the searching process. But in the proposed system, the system would not wait for the entire word to enter; perhaps the searching process starts immediately after the first letter is entered. Because the system performs the

search processes for every letter of a word. This approach considerably increases the speed of the retrieval process.

Tamil language: Tamil is the oldest and truest of the Dravidian speeches. Tamil boasts its literary tradition for more than 2,200 years. Tamil is the most remarkable body of secular poetry extant in India (Tamil is a South Indian language spoken widely in Tamil Nadu in India). Tamil has the longest unbroken literary tradition amongst the Dravidian Languages. Tamil [1] has 12 vowels and 18 consonants. They combined with each other to yield 216 composite characters and 1 special character (aayutha ezhuthu). Summing up there are 247 letters in standard Tamil alphabet.

Tamil literature: It has a rich and long literary tradition spanning more than two thousand years. Contributors to the Tamil literature are mainly from Tamil people from Tamil Nadu, Sri Lankan Tamils from Sri Lanka and from Tamil Diasporas. A revival of Tamil literature took place from the late nineteenth century when works of religious and philosophical nature were written in a style that made it easier for the common people to enjoy. Nationalist poets began to utilize the power of poetry in influencing the masses. With the growth of literacy, Tamil prose began to blossom and mature.

Literature Survey

Language occupies an important position in the history of Indian cultural traditions. The central Government declared Tamil as one of the classical languages on September 17, 2004. Information retrieval [2] of Tamil literature is a difficult work to do because it was used in olden period Tamil format and it was on poetry format as well. Generally morphology approaches [3] are used for information retrieval of Tamil documents (Anand Kumar M), 2009. An IR system returns a list of long documents to a user query. The construction and use of exploration models and search indices consumes processing time, memory, and disk space. Furthermore, in real systems any search and exploration methods must be computationally efficient. In particular, the delay perceived by the users is critical.

It is therefore important to develop methods that can speed up the search process while maintaining high perceived quality, particularly in the range of high precision and low recall which is most crucial in actual user settings. The TAR system is suitable for performing information retrieval using Tamil phonemes and vector space model that were organized for exploration of Tamil document collections

Tamil Phonemes

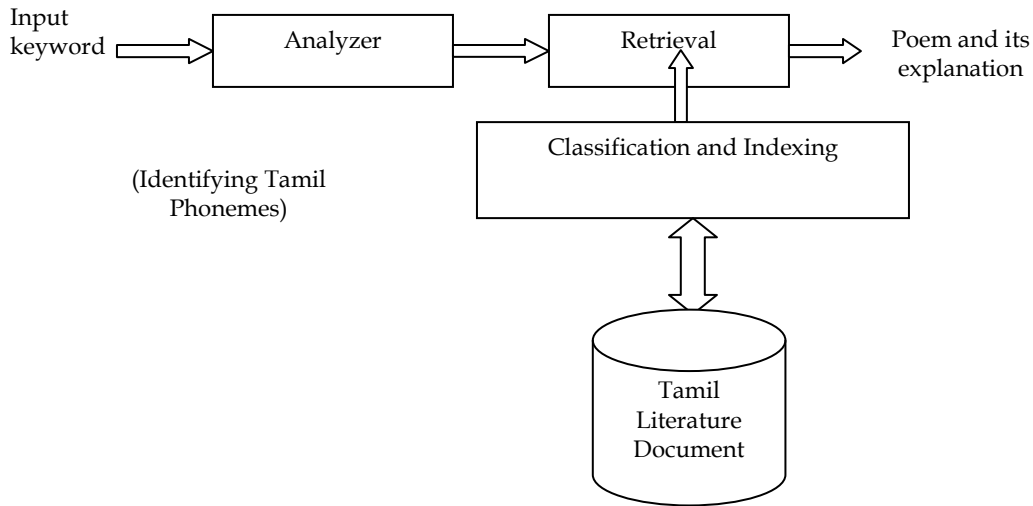
Native grammarians classify Tamil phonemes into vowels, consonants, and a "secondary character", the *āytam*.

Vowels: There are 12 vowels in Tamil, called *uyireluttu* (*uyir* – life, *eluttu* – letter). These vowels are classified into short (*kuril*) and long (*Nedil*), five of each type and two diphthongs, /ai/ and /au/, and three "shortened" (*kurriyl*) vowels. The long vowels are about twice as long as the short vowels. The diphthongs are usually pronounced about 1.5 times as long as the short vowels.

Consonants: Consonants are known as *meyyeluttu* (*mey* – body, *eluttu* – letters) in Tamil. It is classified into three categories with six in each category: *vallinam* (hard), *mellinam* (soft or Nasal) and *itayinam* (medium).

System Architecture

In the TAR system, 4 phases are used. They are (i) Classification (ii) Analyzer (iii) Retrieval and (iv) Spell Checker.



Classification: Most of the ancient Tamil literatures are rendered in the form of poetries. The critical edition of ancient Tamil works include 41 works namely 1) Thirukkural 2) Pura naanooru 3) Aga naanooru 4) Silapathigaram 5) Seevaga chinathamani 6) Manimegalai 7) Kundalakesi 8) Valayapathi 9) Padhinen Mel kanakku (18 Upper Classics) 10) Padhinen Keezh kanakku (18 Lower Classics) etc. Since most of the Ancient Tamil works are in poetry and in anthology forms, a Main class is derived with 41 categories [10] and each category has various sub divisions. The sub divisions are classified as Main topics, Subtitles and Keywords.

Analyzer: Every letter has been analyzed by the analyzer using Tamil phonology[2]. If a key is pressed through the Tamil virtual keyboard, the analyzer would identify the letter through any one of the following - Vowels(V) or Consonants(C). Vowels are again classified into two types, Kuril(k) and Nedil(n). Similarly consonants are classified into three classes with 6 in each class and are called Vallinam(v), Idaiyinam(i) and Mellinam(m). Once the letter is identified, it locates the letter from the Database. The cycle would continue until all the letters in the word are processed. The index position should change for every letter until the complete word is entered.

Eg: புது :- $\text{ப} \rightarrow \text{ப} + \text{உ}$, $\text{து} \rightarrow \text{த} + \text{அ}$, $\text{து} \rightarrow \text{து}$
 (C-v) + (V-k) (C-v) + (V-k) (C-m)

Retrieval: In the retrieval module [12], once the word is found in indices, it retrieves the poem and its explanations from the documents using vector space model [3]. The performance of the search are increased by splitting up the category by 3 ways namely (i) Main topics searches (ii) Subtitles searches

and (iii) Keyword searches. For example, if the input is given from “Main topics”, then the subtitle and keyword could be skipped.

Spell Checker: The spell checker basically checks the spelling of the given word. If the word is wrong or the word is not in the DB, then it would suggest some other word related to the input. Generally the user cannot enter the first letter wrongly, but he may do mid of the word for the letters like

1) ர ற 2) ல ள ழ 3) ன ண ந

So the spell checker is designed in such a way that the first letter would be skipped and the rest of the letters in the word should undergo the spell-check. Basically it skips the first letter of the word and check the rest of the letters of the word. Finally it gets the word from the DB and replaces it. If there is more than one combination, it would list all possible combinations to select. Look at these examples below:

Case: 1 பயண் → பயன்; அரிதல் → அறிதல்

Case: 2 புகல் , புகள் → புகழ் Case: 3 தளை → தலை, தழை

Vector Space Model

In the vector space model [11], we represent documents as vectors. Term weighting is an important aspect of modern text retrieval systems. Terms are words, phrases, or any other indexing units used to identify the contents of a text. Since different terms have different importance in a text, an important indicator – the term weight[8]- is associated with every term. The retrieval performance of the information retrieval system is largely dependent on similarity measures. Furthermore, a term weighting scheme plays an important role for the similarity measure. There are three components in a weighting scheme is

$$a_{ij} = g_i * t_{ij} * d_j$$

Where g_i is the global weight of the i th term, t_{ij} is the local weight of the i th term in the j th document, d_j is the normalization factor for the j th document. Usually the two main components that affect the importance of a term in a text are the term frequency factor[13] (tf), the inverse document frequency factor.

TFIDF weighting: TFIDF is the most common weighting method used to describe documents in the Vector Space Model[12], particularly in IR problems. we assign to each term in a document a weight for that term, that depends on the number of occurrences of the term in the document. In addition, IDF measures how infrequent a word is in the collection. This value is estimated using the whole training text collection at hand. The simplest approach is to assign the weight to be equal to the number of occurrences of term t in document d .

$$\text{weight}_{t,d} = \log(tf_{t,d} + 1) \log \frac{n}{x_t}$$

where $tf_{t,d}$ is the frequency of word t in document d , n is the number of documents in the text collection and x_t is the number of documents where word t occurs. Normalization to unit length is generally applied to the resulting vectors

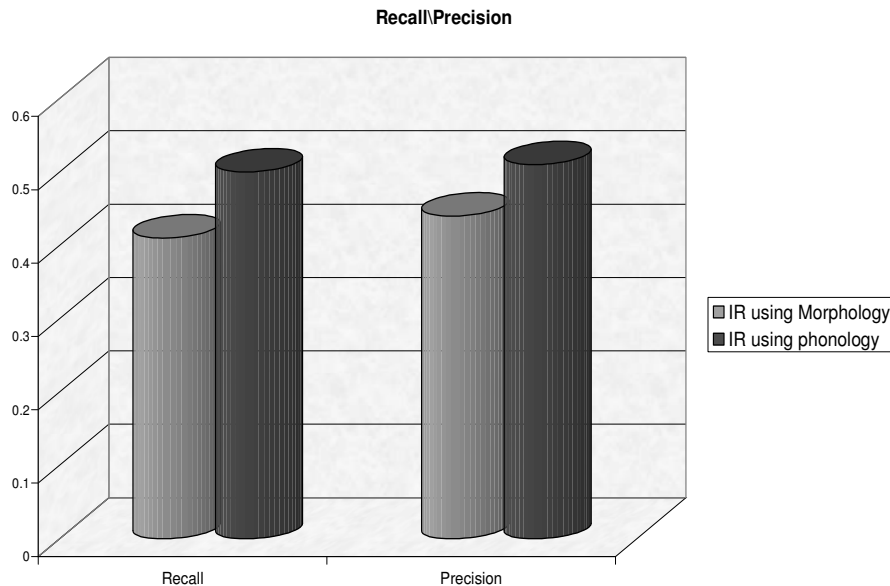
Evaluation

Objective evaluation of search effectiveness has been a cornerstone of IR. Progress in the field critically depends upon experimenting with new ideas and evaluating the effects of these ideas, especially given the experimental nature of the field. Since the early years, it was evident to researchers in the community that objective evaluation of search techniques would play a key role in the field. The two desired properties that have been accepted by the research community for measurements of search effectiveness are *recall*: [3] the proportion of relevant documents retrieved by the system; and *precision*: the proportion of retrieved documents those are relevant.

Experiments Results : Information retrieval systems comparison:

Input word is புகழ்

Title	Recall	Precision
IR using Morphology	0.41	0.44
IR using phonology	0.50	0.51



Conclusion

The TAR system would be pretty much helpful for all Tamil literates and students to search and learn. The TAR system focuses on Tamil phonemes. Also performance tuning has also been done in the system. In addition with that, the TAR system has the spell checker concept which makes the system to search the data reasonably quick. This system currently supports 41 categories alone. In addition to this we can further add more documents in future. However this concept will be designed more useful to the users, in future.

References

1. Abirami.S and D. Manjula , Enabling Intelligent Information Retrieval from Tamil Document Images, *Asian Journal of Information Technology*, 996-1000, 2006
2. Abirami.S and D. Manjula, Feature string-based intelligent information retrieval from Tamil document images, *International Journal of Computer Applications in Technology*, Publication, Volume 35,150-164, 2009
3. Amit Singhal,Google, Modern Information Retrieval: A Brief Overview,*ieee tranactions* , IEEE Computer Society Technical Committee on Data Engineering, 2001
4. Anand kumar M, Dhanalakshmi V, Rajendran S, Soman K P, A Novel Approach to Morphological Analysis for Tamil Language,*Internet Tamil Conference*, 2009
5. Bayardo, R. J., Ma, Y., & Srikant, R. Scaling up all pairs similarity search. In *Proceedings of the 16th international conference on World Wide Web (WWW '07)*, pp. 131-140, New York, 2007
6. Gorman.J, & Curran J. R., Scaling distributional similarity to large corpora. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 361-368, 2006.
7. Guo xian tan, Chsitian vivard, Alex.c, Information Retrieval Model for Online Handwritten Script Identification, *International Conference on Document analysis and Recognition*, Spain, 2009
8. Holger, Billhardt, Victor Maojo, A context vector model for information retrieval, *Journal of the American Society for Information Science and Technology*, Volume 53, Pages: 236 - 249, Year of Publication: 2002
9. Massimo Melucci, A basis for information retrieval in context, *ACM Transactions on Information Systems (TOIS)*, volume.26 n.3, p.1-41, June 2000
10. Rajan.K,Ramalingam.V, M.Ganesh, Automatic classification of Tamil documents using vector space model and artificial neural network, *Expert Systems with Applications: An International Journal*, Volume 36 , Issue 8 Pages: 10914-10918, 2009
11. Ray Larson, Marc Davis. SIMS 202: Information Organization and Retrieval. UC Berkeley SIMS, Lecture 18: Vector Representation, 2002.
12. Wan, V. N. Anh, I. Takigawa, and H. Mamitsuka. Combining vector-space and word-based aspect models for passage retrieval. In E. M. Voorhees and L. P. Buckland, editors, *Proc. 15th Text Retrieval Conference (TREC 2006)*, Special Publication 500-272, November 2006.

Speaking tool in Tamil for vocally disabled

Shashikiran K¹, Abhinava¹, Swapnil Belhe², A G Ramakrishnan¹

¹MILE Lab, Dept of Electrical Engineering, Indian Institute of Science, Bangalore 560 012.

²Gist Group, CDAC Pune.

Shashikiran@mile.ee.iisc.ernet.in, abhinav@mile.ee.iisc.ernet.in,
ramkiage@ee.iisc.ernet.in, swapnilb@cdac.in

Introduction

It has always been a challenge to bridge the gap between vocally disabled and the masses. The development of sign language has only been partially successful in bridging the gap. It requires persons conversing to know the sign language. Our work is a conscious effort to overcoming this pitfall. The proposed methodology is a combination of two different entities namely (Online Handwriting Recognition) OHR and (Text to Speech) TTS. OHR deals with recognition of handwritten words online. In an on-line handwriting recognition system, machine recognizes the writing while the user writes. The OHR output is in Tamil Unicode & becomes the input to TTS. Unicode is a globally accepted encoding format which makes our application viable to be used in various circumstances. TTS is a system that takes Unicode text and produces natural and intelligible speech in that language. This enables the patients, who had laryngectomy and tracheotomy as well as the vocally challenged to communicate effectively. As vocal disability may be congenital or because of ailments like oral or throat cancer, our method serves equally to both.

The tool is based on a hand-held, Tablet PC based on Intel Atom processor. The user can write one sentence at a time on the screen using the stylus and then click the button "Speak". The sentence is recognized and then converted into speech and spoken out. Thus, the patient can call the attention of the nurses or his relatives in another room easily.

Details of the OHR module:

In Online handwriting recognition, a machine recognizes, as a user writes on a pressure sensitive screen with a stylus. The stylus captures information about the position of the pen tip as a sequence of points in time. The sequence of point between a PEN DOWN and PEN UP signal defines a stroke. This spatio-temporal information of the character being traced is the only input available to the online recognition system. Also given a character, one can capture the different writing styles using the information from the stylus.

Given a Tamil word, we first run a segmentation algorithm to identify the individual symbols. This algorithm segments word level data into symbol level data, as the modeling of the data is done at the symbol level. The recognition is performed at this level and results are concatenated to form the words. The extracted symbols are subject to the following preprocessing modules: smoothing to remove noise, resampling to a fixed number of points for speed normalization and size normalization.

Once symbols are brought to a standardized form, a set of seven features namely ,

Pre-processed X & Y co-ordinates:	Preprocessed data points (x,y) are themselves good features
Pen direction angle:	At each sample point, the direction of pen tip movement from that point to the next point can be used as a feature
Normalized first derivatives of X & Y:	Derived at each sample point of the preprocessed Tamil symbol, are also used.
Normalized Second derivatives of X & Y:	Same as above.

The preprocessing techniques & features are discussed elaborately in Rituraj et al [2]. These features are then fed to the SDTW classifier for recognition of the Tamil symbol.

Statistical Dynamic Time warping (SDTW): In SDTW, a reference character is represented by a sequence $Q=(Q1;Q2;Q3;Qlq)$ of statistical quantities (states), as shown in Fig 1.

These statistical quantities include

1. Discrete probabilities that statistically model the transitions between states. We have empirically used 20 states in our work.
2. A continuous probability density function that models the feature distribution at each state.

We have modeled this distribution as a multivariate Gaussian distribution for each of the 20 states.

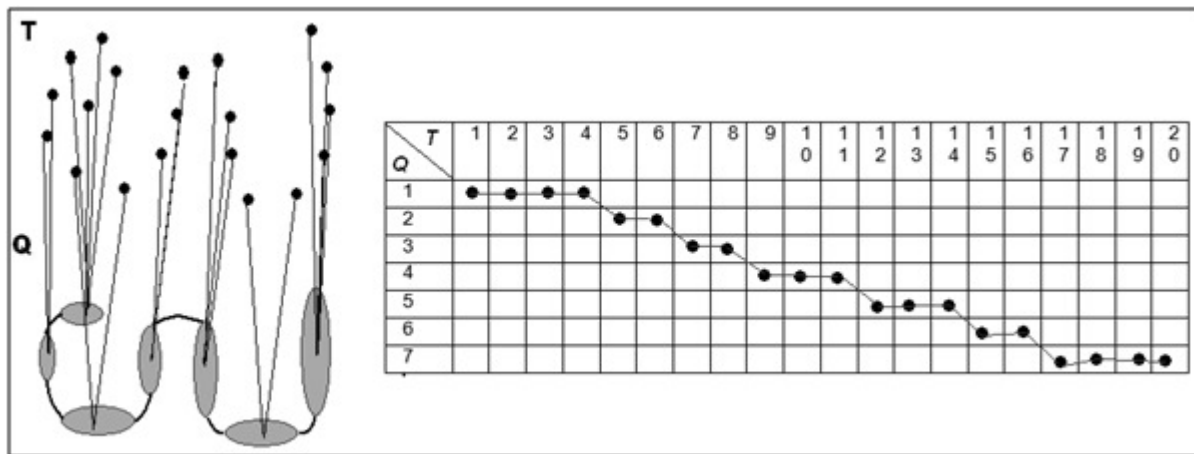


Fig 1. Transitions between states in SDTW

While testing, the SDTW distance of test pattern to the reference model of each class is computed. and the test pattern is assigned the label of the class giving minimum SDTW distance. The definition of SDTW distance is different from that of DTW and is given in Claus et al [1].Fig.1 shows how the matching takes

place between the reference model and Test pattern. Matrix in Fig.1 shows the DTW path (path of best SDTW matching).

SDTW distance is the negative log state optimized likelihood of pattern T generated by the model Q, with optimal state sequence S, given by the Viterbi algorithm. So, models in SDTW frame work are similar to HMMs of particular type with state prior probabilities $\pi=(100\dots0)T$ and are left to right models with step size of at most 1 and with null transitions (transitions that allow change in state without observation change i.e. transitions (0,1)). So the models in SDTW frame work can be trained by algorithms used for training HMMs. In our work, we used segmental K-means algorithm [3] for training SDTW model parameters.

Unicode Generation: Based on rules derived from the language, valid Unicode string is generated from the output labels of the classifier. This string is the input for the TTS module.

Details of the TTS module

The TTS is based on concatenative waveform synthesis [4]. There are 1026 phonetically rich sentences spoken by a professional Tamil speaker, which has been segmented and annotated at the phone level. The input text is passed through a grapheme to phoneme conversion module, after text normalization. Certain pause rules are added based on a preliminary POS tagger and several rules. The phonemic text is parsed into the basic units for concatenation and the units are searched for, from the synthesis speech database, based on context and prosodic parameters. The selected speech units are concatenated in the waveform domain using a pitch continuity metric [5] and a pitch synchronous concatenation methodology [6]. It takes directly Unicode Tamil text as the input and produces a .wav file as the output. This wav file is directly played by the tablet PC and hence, when the user writes a word, it will be read after a second by the system.

References:

1. Clauss Bahlmann and Hans Burkhardt , "The writer independent online handwriting recognition system Frog On Hand and cluster generative statistical dynamic time warping", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.26, No. 3, pp 299-310, March 2004.
2. Rituraj Kunwar et al. A HMM Based Online Tamil Word Recognizer, 8thTamil Internet Conference, 2009,University of Cologne , Germany.
3. Rakesh Dugad and U. B. Desai "A tutorial on hidden markov models",SPANN- 96.1 May 1996.
4. K Parthasarathy, "A research bed for unit selection based text to speech synthesis system", MS thesis, Dept. of electrical Engineering, IISc Bangalore, Feb. 2009.
5. Vikram Ramesh Lakkavalli, Arulmozhi P and A G Ramakrishnan, "Continuity Metric for Unit Selection based Text-to-Speech Synthesis", accepted for oral presentation at SPCOM 2010, Bangalore, July 23-26.
6. Vikram Ramesh Lakkavalli and Ramakrishnan AG, "A Novel Method of Epoch Extraction for Concatenative Text-To-Speech Speech Synthesis" submitted to Interspeech 2010, Japan, Sept. 26-30, 2010.

Standardisation of Modern Standard Tamil Transcription for Computational Tamil

Dr. Punal K Murugaiyan
Prof. Rtd. Annamalai University
Tamil Nadu, India.

SECTION 1 - Antiquity

Tamil is the senior most and uniquely distinguished member of the Dravidian family of languages, possessing maximum possible cognates and other common features of the group. Some scholars of academic reputation both native and foreign believe that it has at least sources to interpret, nucleus elements to share with quite a few languages of the world. Its phonology itself stands witness for its antiquity and primitiveness. Markedness features of the phonological pattern are the characteristic indices which could vouch for the earliest evolution i.e. the primitiveness of a language. Unmarked phonological features are the simple, earliest and most primitive ones. All the phonological elements i.e. phonemes of the Tamil language, without exception are surprisingly unmarked. Unmarked features, since simple and designating single articulation only, naturally the sounds unmarked must have developed at the initial stage of the evolution of speaking. The vowels and consonants of the Tamil language are unmarked for any of the phonological features of articulation, since the alphabets whether invented, adopted or adapted, as Alexander John Ellis (1842, p.2) points out might have been representing, at the time of introduction, only one individual sound rather only one allophone. Even now, though a lot of phonetic variants had emerged, the unmarked characteristics of obstruent and sonorant remain unchanged. This kind of unique unmarked phonological feature has not been so far noticed in any of the world languages except the one, Curier, the aboriginal Australian language, which again reinforces the theory of Lemurian vast land.

Standard writing

As Edward Jewitt Robinson (Tamil Wisdom, 1873, p.3) observes, the Tamil Sangams were like “..... the Royal Academy of Sciences founded by Louis XIV at Paris and made it a rule that every literary production should be submitted to their Senatus Academicus before it was allowed to circulate in the country, for the purpose of providing the purity and integrity of the language”. Even now the standard form of witing was strictly adhered while the pronunciation, the actual torch bearer, the communication base had/ has never been insisted, not even for name sake, cared for. The mushroom growth of the Tamil dictionaries, due to business motive, never spare even little time on the entry of pronunciation. Even now, the supposed edition of the great Tamil Lexicon of the University of Madras has ruled out the phonetic transcription, hammering it out mercilessly. It is very painful to notice that the scholarly comment of P.S. Subrahmanya Sastri on the entry of āytam in the existing volume remains uncared (History of Grammatical Theories in Tamil, 1934, p.68).

The said entry reads, Sastri quotes translating, “..... the 13th letter of the Tamil alphabet occurring only after a short initial letter and before a hard consonant as *aLkam*, and pronounced sometimes as vowel and sometimes as consonant”. But the medieval authors of grammatical treatises themselves are clearly stating that it is neither a vowel nor a consonant, and fix its pronunciation as guttural fricative which coincides with the current one. The irony is one of the advisers of the new edition, may be going to appear ingeniously discovered that *āytam* is a voiced double implosives!

Standard pronunciation

In the Sanskrit grammatical literatures there are separate treatises on phonetics like *sikshas* (general phonetics) and *pratisakyas* (phonetics of particular school of veda). Pronunciation is equally insisted along with the other features of grammatical principles. It is said that he who knows the distinction between the length and tone can secure a seat by the side of the master. Such an importance has been given for correct speaking. Due to the advent of linguistic science, pronunciation dictionaries even on specialized subjects appeared in European and a few other languages and phonetic transcription even in the regular dictionaries too, finds entry.

Importance can't be ignored

It is very unfortunate that neither the grammarians nor the scholars of the other related fields of the Tamil language up till now have taken cognizance of this subject. But that can't be no longer remain discarded as the demand is waiting right at their door steps due to unavoidable need of the monstrous advent of the computer pervasion in every walks of modern life. So no longer the Tamil savants ignore the importance of the pronunciation features of speech of the Tamil language as it had already established its supremacy in other languages. No one can think of an alternative for computer speech of Natural Language Processing

Standardization of transcription

Tamil since holding an extraordinary situation than the rest of the languages of the world i.e. being in current use, since the beginning of the historical period and preserving the ancient and primitive features of speaking due to its conservative elders, is now desperately in need of alteration in its alphabet giving space to new sounds which entered due to importation of knowledge and culture. Even the earliest extant grammar *Tolkappiyam* itself gives way to annexing new sounds, suggesting rules for accommodation. But in spite of the grammatical sanctioning the so called purists wage pseudo war for fancy sake.

New entries

As seen in the foregone paragraph one cannot go but accommodating new words into the corpora admitting new sounds. As it was already happened to be the practice of using Grantha letters for the Sanskrit and Prakrit borrowings and appended them to the existing alphabet, if any new sounds added may be represented by the Grantha letters which go in harmony with the Tamil scripts since having similar shape and structure. If one thinks of diacritic forms-allograph that spoil the whole beauty of the structure and system of the alphabet.

Standard variety

First of all, the standard variety of the speech, the pronunciation must be chosen as not only being scientific but also wildly to be acceptable and practicable. In this connection the elites' and well accepted speech, mostly used in formal occasions of the educated people has to be taken as the data, ignoring all voice quality, intonation and idiolect variations; mostly isolated words in all length are taken in to consideration.

Phonemes

There are 10 vowels (5 short and 5 long), 25 consonants (18 native: 6 plosives, 6 nasals and 6 approximants; and 7 marginal: 4 voiced plosives and 3 fricatives) – they are in the native and Grantha forms of script.

Ex: அ, ஆ; இ, ஈ; உ, ஊ; எ, ஏ; ஒ, ஓ
 a, ā; i, ī; u, ū; e, ē; o, o:

Native form

க, ச, ட, த், த், ப்

K, t̪, t̪, t̪, p

ங், ஞ், ண், ந், ம், ன்

ᱠ, ᱡ, ᱢ, ᱣ, ᱤ, ᱦ

ய், ர், ல், ள், ழ், வ்

J, r, l, ʎ, ɟ, v

Grantha form

𑌀, 𑌁, 𑌂, 𑌃

g, ʈ, ɖ, b

ṣ, ś, ḥ

s, ʃ, h

Allophones or phonetic variants

To avoid the unwieldy length which may overrun the present article, the acoustic allophones which are though more important and crucial when this present work of interest is concerned are deliberately avoided and only the articulatory allophones being more simple and handy to describe and explain are taken as the main subject of discussion.

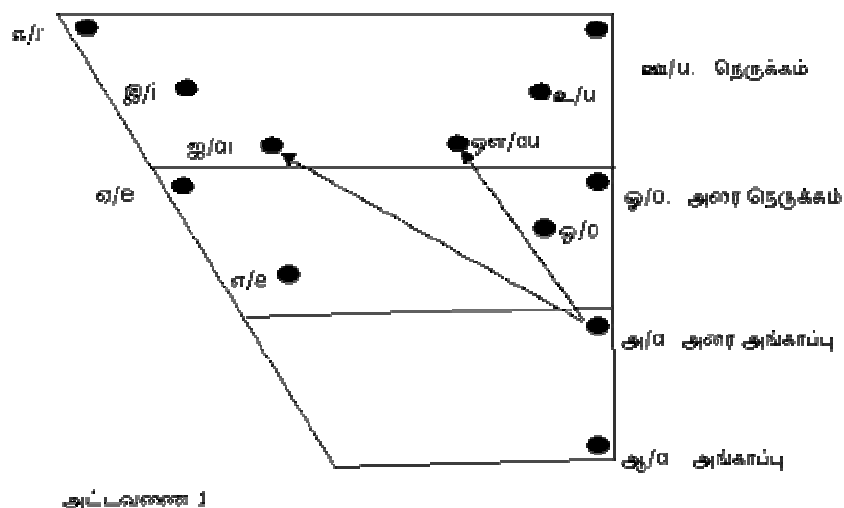
There are altogether 91 allophone: 52 being vocoids and 39 being contoids.

SECTION 2 - Vowel and consonant phonemic charts

உயிரொலியன்கள்

உயிரொலியன்கள், மெய்யொலியன்கள் ஆகிய வரிவடிவங்களுக்கு ஒப்பான ஞா. ஒ. நெ. வரிவடிவங்கள் கண்டறியப்பட்டு அவற்றின் வரன்முறைக்கு ஏற்ப நிரல்படுத்தி அட்டவணையில் அமைக்கப்பட்டுள்ளன. உயிரொலியன்கள் வாயறையின் அங்காப்பு, நாவின் நீட்சி, திரண்டமைவு ஆகியவற்றின் அடிப்படையில் ஞா. ஒ. நெ. நாற்கர அட்டவணையில் இடம்படுத்தப்பட்டுள்ளன.

அட்டவணை ஒன்றில் உள்ள படத்தில் உயிரொலியன்களின் இடங்கள் வாயறையில், தாடைகளின் அங்காப்பு, நாவின் திரண்டஅமைவுக்கு ஏற்ப அடையாளமிட்டுக் காட்டப்பட்டுள்ளன. இவ்விடங்கள் ஒவ்வொன்றும் அந்தந்த உயிரொலியன்களின் மொத்தப் பிறப்பிடத்தைக் காட்டிநிற்பன. அதாவது,



அட்டவணை 1

உயிரொலியன்களின் மாற்றொலிகள் இந்த இடங்களிலேயே பிறப்பவை ஆகும்.

அட்டவணை 1.அ-வில் உயிர் ஒலியன்கள் பிறக்கும் இடங்களை எளிதில் புரிந்துகொள்ளும் வகையில் வாயறையின் நீண்ட அமைப்பினையும் (முன், நடு, பின்) உயர்வு அமைப்பினையும் (மேல், இடை, கீழ்) கொண்டு நிரல்படுத்திக் காட்டப்பட்டுள்ளன.

	முன்	நடு	பின்
நெருக்கம்	இ, ஈ i, i:		உ, ஊ u, u:
அரை நெருக்கம்	எ, ஏ e, e:	ஐ, ஓ ai, oi	ஓ, ஔ o, o:
அங்காப்பு		அ, ஆ a, a:	

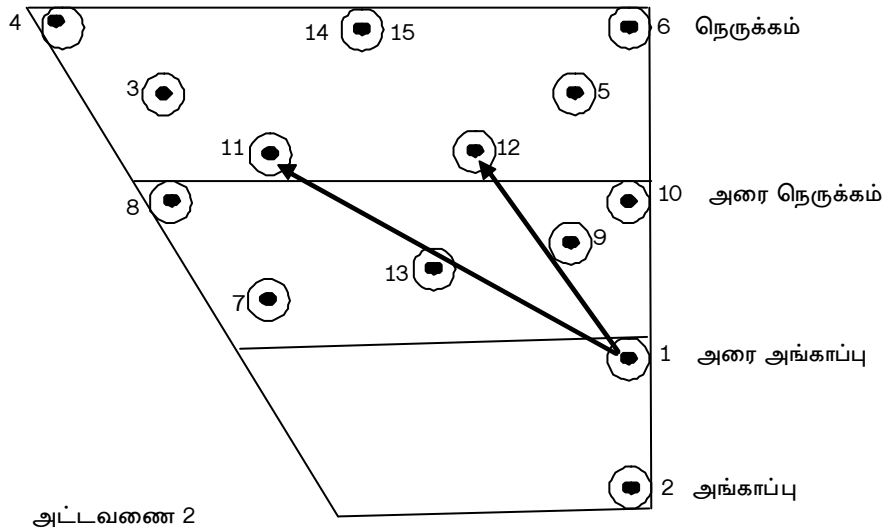
மெய்யொலியன்கள்

	கடை அண்ணம்	இடை அண்ணம்	முன் அண்ணம் / வளைநா	அண்பல்	பல்	இதழ்
அடைப்பொலி	க் k		ட் t	ற் t	த் t	ப் p
அடைப்புரளி		ச் ச				
மூங்கொலி	ங் ṅ	ஞ் ṅ	ண் ṅ	ன் n	ந் ṅ	ம் m
அதிரொலி				ர் r		
மறங்கொலி			ள் l	ல் l		
உயிர் - பொலி		ய் y	ய் y			வ் v

மெய்யொலியன்கள் கிடைவரிசையில் ஒலியுறுப்புகளும் நெட்டு வரிசையில் முயற்சியும் நிரல் படுத்தப்பட்டு ஒலியுறுப்புக்கள், அவற்றிற்கு எதிரான முயற்சி ஆகியவற்றின் அடிப்படையில், குறிப்பிடப்பட்ட ஒலியுறுப்பாலும் முயற்சியாலும் விளைவிக்கப்படும் ஒலியனின் வரிவடிவம் ஞா. ஒ. நெ. அட்டவணையில் கண்டவாறு இடம்படுத்தப்பட்டுள்ளது (அட்டவணை எண் 3).

உயிர் மாற்றொலிகள்

தமிழ்மொழியின் உயிரொலிகளில் பொதுவான இயல் புத் தன்மை உள்ளது. i. மொழிமுதல் உயிரொலிகள் அவற்றுடன் ஒத்த உராய்வு நிலையை அதாவது முன்உடம்படு மெய் (onglide) பெற்றுவரல்; ii. வளைநா ஒலியை முந்துறும் பொழுது வளைநா சாயல்பெற்றுவரல். iii. உகரக்குறில் மொழிக்கு முதலில் உகரம் உள்ள அசையை அடுத்து வரும்பொழுதும் இடங்குவிந்து வரல், ஏனைய இடங்களில் பரந்த இதழுடன் வரல்.



அட்டவணை 2இல் உள்ள நாற்கரத்தில் 12 உயிரொலியன்களின் பிறப்பிடங்களாக 15 இடங்கள் காட்டப்பட்டுள்ளன. இப்பதினைந்து இடங்கள் குறிப்பிட்ட உயிரொலியனின் மாற்றொலிகளின் பிறப்பிடங்களாகும். எண்கள் அவ்வவ் இடங்களில் பிறக்கும் மாற்றொலிகளைக் காட்டுவதற்காக அமைக்கப்பட்டுள்ளன.

	முன்	நடு	பின்
நெருக்கம்	4, 5	14, 15 1-11, 1-12	6, 5
அரை நெருக்கம்	6, 7	15	10, 9
அரை அங்காப்பு			1 1-11, 1-12
அங்காப்பு			2

அட்டவணை 2.அ-இல் காட்டப்பட்டுள்ள உயிர் மாற்றொலிகளின் 15 பிறப்பிடங்களும் எளிமை கருதி நிரல் படுத்தப்பட்டுள்ளன.

- 1: [ʌ]; [ʌ̃]; [ɔ]; [ʔʌ]; [ʔʌ̃]
13: [ə]
2: [ɑ]; [ɑ̃]; [ʔɑ]; [ʔɑ̃]
3: [ɪ]; [ɪ̃]; [ɪ̃]; [ɪ̃]; [ɪ̃]; [ɪ̃]
4: [i]; [ĩ]; [ĩ]; [ĩ]; [ĩ]
5: [ʊ]; [ʊ̃]; [ʊ̃]; [ʊ̃]
14: [ɨ]; [ɨ̃]
15: [ɘ]; [ɘ̃]
6: [u]; [ũ]; [ʊu]; [ʊũ]; [ũ]
7: [ɛ]; [ɛ̃]; [ɛ̃]; [ɛ̃]
8: [e]; [ẽ]; [j̃e]; [j̃ẽ]; [ẽ]
9: [ɔ]; [ɔ̃]; [ʊɔ]; [ʊɔ̃]
10: [o]; [õ]; [ʊo]; [ʊõ]; [õ]
11: [ʌ̃]
12: [ʌ̃~ ʌ̃]

மெய் மாற்றொலிகள்

i) குரலில் அடைப்பொலி

- கடல் (kஅட்அல்), அக்கம், (அkkஅம்),
வெட்கம் (வ்ள[kஅம்]
சிவப்பு (ஶ்இவ்அப்ப்உ), அச்சம் (அஶ்ஶ்அம்),
கட்சி (க்அஶ்ச்இ)
அட்டன் (அட்டஅன்), வெட்கம் (வ்ள[kஅம்]
மாற்றார் (ம்ஆஶ்ஶ்ஆர்), கற்பு (க்அஶ்புஉ)
தனி (த்அள்இ), சத்தம் (ச்அஶ்அம்), சக்தி (ச்அக்இ)

பலி (pஅல்இ), அப்பர், (அppஅர்),
நட்டி (ந்அ[ரஉ)

ii) குரல் அடைப்பொலி

தங்கம் (த்ங்ஐஅம்)
தஞ்சம் (த்அஞ்சுஅம்)
தண்டம் (த்அண்டஅம்)
தந்தம் (த்அந்தஅம்)
தம்பம் (த்அம்பஅம்)
கன்று (க்அன்றுஉ)

இயங்கும் உறுப்பு		கடை நா	இடை நா	முன் நா, நுனி	நுனி நா	நா நுனி	கீழ் இகழ்
நிலை முயற்சி	கடை அண்ணம்	அண்ணச்சாயல்	இடை அண்ணம்	முன் அண்ணம்	அண் பல்	பல்	மேல் இகழ்
அடைப்பு	k g	k' g'	c	t d	t	t d	p b
வருடல்				r			
அதிர்வு விடுப்பு					t' d'		
அடைப்புரலை			ஶ ஐ				
மூக்கினம்	η	η'	ɳ	ɳ	η	η	m
வருடல்				ɳ'			
உரலை	x ʏ		ç		s	ð	β
அதிர்வு					r		
வருடல்					r		
மருங்கு				l		l	
வருடல்				l'			
உயிர்-போலி			j				u
மருங்கு சாயல்				ɹ			

iii) உரசொலிகள்

அகம் (அXஅம்)
அசல் (அSஅல்)
காது (க்ஆ[உ)
சுபம் (ச்உ[அம்)

iv) வருடொலி

கடை (க்அஐ)

சாறு (ச்ஆஉ)

மெல்லின மாற்றொலிகள்

தங்கம் (த்நக்அம்)

மஞ்சம் (ம்அநிச்அம்)

பாண்டம் (ப்ஆநட்அம்)

மணம் (ம்அந்அம்)

மன்றம் (ம்அந்அம்)

எந்தை (எந் தை)

நகரம் (ந அக்அர்அம்)

கம்பம் (க்அமப்அம்)

இடையின மாற்றொலிகள்

பயன் (ப்அ ிஅன்),

பையன் (ப்அய்ய்அன்)

பரம் (ப்அ ரஅம்)

பயிர் (ப்அய்ய்இர)

வலம் (வ்அலஅம்)

அவன் (அ ஸஅன்)

அவ்வை (அவ் ஸுஐ)

பழம் (ப்அ டஅம்)

வளம் (வ்அ ல்அம்)

கள் (க்அல்)

ஆய்தம்

கல் + தீது > கற்றீது > கஃறீது;

முள் + தீது > முட்டீது > முஃட்டீது;

அவ் + து > அத்து > அஃத்து;

இவ் + து > இத்து > இஃத்து;

பல் + து பத்து பஃது போன்ற சொற்களின் அமைப்பு முறையினை வரலாற்று அடிப்படையில் ஆயப் புகின், தெளிவு பெறமுடியும். மேலே கண்ட சொற்களில் **வ், ள், வ்** போன்றவற்றின் திரிபே ஆய்தம் என்பது புணர்ச்சி இலக்கணத்தால் அறியக் கிடைக்கும்.

Realization of Tamil Syllables Text To Speech Transferring System Using FPGA

T.Jayasankar
Lecturer/ECE
Anna University Trichy
jayasankar@tau.edu.in

Dr.J.Arputha Vijaya Selvi
Professor
Kings College of Engg
arputha09@gmail.com

Prof .R. Rajendran
Tamil University
Pudukkottai Thanjavur
raj_ushus@yahoo.com

Abstract

Design and development of the functional architecture of Tamil TTS engine using Field Programmable Gate Array (FPGA) is reported. It is a stand alone hardware system without any operating system (OS) to sense, identify and convert the Tamil mono-syllable text to speech output which is important for the visually impaired persons. The proposed model in FPGAs optimises the parameter extraction to perform efficient speech synthesis.

Keywords – Tamil text to speech, FPGA.

1. Introduction

Text to Speech (TTS) Synthesis is an automated encoding process which converts a sequence of symbols (text) conveying linguistic information, in to an acoustic waveform (speech). In recent years, text to speech synthesis technologies for different language are growing rapidly.

Speech synthesis based on syllables seems to be a good possibility to enhance the quality of synthesised speech. As far as the production of speech is concerned, syllable is the minimum possible speech segment which can be spoken in isolation.

A Tamil voice system is expected to pave way for developing creative applications to enable users to hear Tamil content in voice form. The key concept behind this voice engine is Text-To-Speech conversion. This conversion uses the method of concatenating the syllables to generate the required words. A system used for this purpose is called a speech synthesizer and can be implemented in software or hardware. Hardware implementation of the above system is achieved by FPGA [Cyclone -II], and Verilog HDL and synthesizer tools. Minimal hardware makes the system to achieve cost effective, compact, very less power consumption and speed.

Introduction to Tamil Languages

Over 65 million people worldwide speak Tamil, the official language of the southern state of Tamil Nadu, and also of Singapore, Sri Lanka and Mauritius. It will be a boom to Tamilian, if the user interface with the system is in Tamil, that too if it is in the form of speech.

Nature Of Tamil Scripts: The basic units of the writing system in Indian languages are characters which are an orthographic representation of speech sounds. A character in Indian language scripts is close to a syllable and can be typically of the form: C*VC*, where C is a consonant and V is a vowel. There is fairly

good correspondence between what is written and what is spoken. Typically there are about 35 consonant and 18 vowel characters. However, in Tamil there are fewer characters than many of the other Indian languages. There are 13 vowels and 18 consonants characters. Some of the consonants have more than one pronunciation and in effect there are 41 phones.

Creation of Tamil Speech Database

To build a unit selection voice, typically a small set of letter is recorded by a native Tamil speaker in a recording studio. The speaker uttered the sentences into a stand mounted microphone placed in front of her or his. The speech data was recorded at 44 KHz, mono channel at 16 bits per sample. After the recording it was down sampled to 16 KHz for further processing.

Experimental Setup

The module used the previously stored phonetics sound to reproduce the word. The actual process of this project is by pressing a key in the PS/2 keyboard which is connected to the ALTERA DE1 board, a particular Tamil syllable which is pre-recorded in the SD card should be play backed.

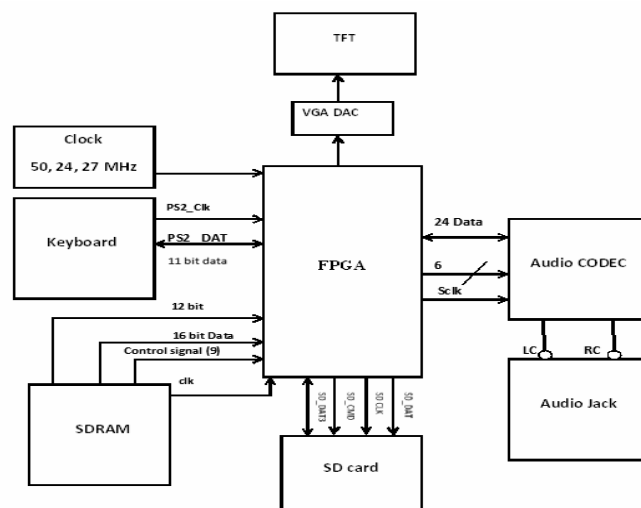


Fig.1. Data flow in TTS system

Here 2 tools are used,

1. Quartus II tool and,
2. NIOS II IDE.

In the Quartus II tool, by using SOPC builder, the hardware components and their connections are created. By using NIOS, the software coding is developed. SD CARD acts as the pre-recorded component (recorded format of Tamil syllables).

SD card is inserted into the SD card driver of the DE1 board and the protocol used here is SP1 mode. When the program is started, the contents of SD card are moved to the FPGA and from there it is sent to

audio codec through I2C protocol bus. On the audio codec the Tamil syllables gets processed and play backed.

Input to the system is the scan code of the keyboard which is connected through PS/2 connector of the DE1. By pressing a key, for e.g., by pressing V the scan code of 2A gets compared with the corresponding recorded Tamil syllables. For V the syllable of “va” will be heard and gets displayed.

If the incoming scan code has no recorded syllable means, it will play back the default values. (i.e.,) the scan code gets compared and the particular memory content is played.

Description and Design of the TTS Transferring System

KEYBOARD: The PS2 port is a widely supported interface for a keyboard and mouse to communicate with the host. The PS2 port contains two wires for communication purposes. One wire is for data, which is transmitted in a serial stream. The other wire is for the clock information, which specifies when the data is valid and can be retrieved. The information is transmitted as an 11-bit "packet" that contains a start bit (Logic 0) followed by 8 data bits (LSB First), one odd parity bit (odd parity), and a stop bit (Logic 1). Each bit should be read on the falling edge of the clock.

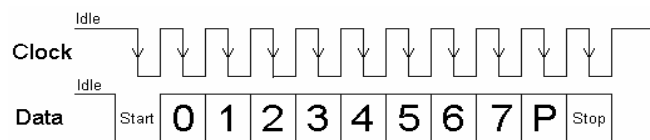


Figure. 2 Timing diagram of a PS2 port

The above waveform represents a one byte transmission from the Keyboard. The keyboard may not generally change its data line on the rising edge of the clock as shown in the diagram. The data line only has to be valid on the falling edge of the clock. The Keyboard will generate the clock. The frequency of the clock signal typically ranges from 20 to 30 KHZ. The Least Significant Bit is always sent first.

The system hardware architecture is shown in figure 3 including CPU, UART, tri-state Bridge, ram and I/O controls, which are all reusable. Such design methods not only makes it Modulization, but also greatly reduce the design cycle of the system.

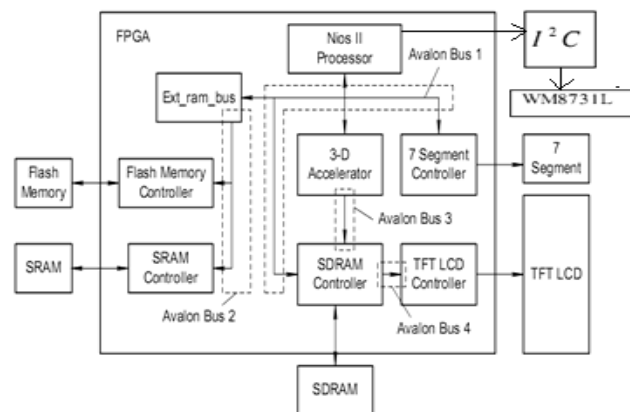


Fig. 3. Hardware architecture

NIOS II SOFTCORE PROCESSOR: Nios II is a high performance 32-bit softcore processor. The processor is configured on an Altera Cyclone II FPGA. Custom instructions are added to improve system performance; furthermore, more on-chip rams can be added to improve data processing capacity.

SD card: Many applications use a large external storage device, such as a SD card or CF card, to store data. The DE1 board provides the hardware and software needed for SD card access. The size of the SD card should be less or equal to 2GB. Also, it is required to be formatted as FAT (FAT16 or FAT 32) File System in advance. The system requires a 50 MHz clock provided from the board. The SD 1-bit protocol and FAT File System function are all implemented by NIOS II software. The software is stored in the on-board SDRAM memory.

SDRAM: In order to store the reference, the 512 kB SRAM module built in the board is used. There are three memory modules on the Altera DE2: a 4 MB Flash memory chip, an 8 MB SDRAM chip and a 512 kB SRAM chip. While the Flash module provides a vast amount of non-volatile storage, it is very slow with respect to the main system clock. The SDRAM chip is very fast and has a very large storage capacity, but it require a very sophisticated controller to be operated. This makes the SRAM chip an obvious choice. Even though it is not the fastest nor the largest, it has ten times the required storage capacity needed for this project, and it is fast enough (since it can perform a read or write operation in less than 20 ns, i.e. a system clock period) so as to avoid any timing issues. Moreover, it is a fairly simple device and can easily control.

AUDIO CODEC: In this project the CODEC is used as both test equipment for other modules and as secondary input/output for the audio system. The DE1 board provides high-quality 24-bit audio via the Wolfson WM8731 audio CODEC. The WM8731 is controlled by a serial I2C bus interface, which is connected to pins on the Cyclone II FPGA. Initialization of the CODEC is through a standard I²C (Inter-Integrated Circuit) bus and sound transfer is through a 3-wire bus, which in this project is defined as a standard I^S (Inter-IC Sound) bus.

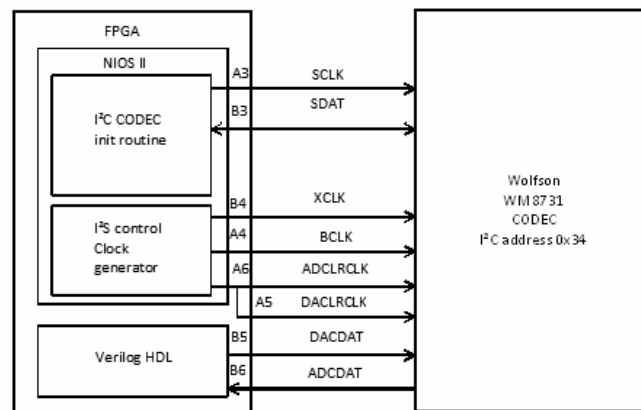


Figure 4. Connection diagram of CODEC part of the DE1 board

The figure 4 above shows the structure of the CODEC connection to the control numbers on the CODEC are generated using the NIOS II processor. This chip supports microphone-in, line-in, and line-out ports, with a sample rate adjustable from 8 kHz to 96 kHz. WM8731 contains A/D, D/A modules with a high

sample rate and quantization precision. We will use 8kHz sample rate and 16 bit quantization precision in this design.

In voice acquisition part, since A / D is the serial data output, a serial to parallel data conversion and control of the SRAM Verilog module is needed. Voice report is communicated with CPU via GPIO; different voice is played according to different verification result. GPIO control is done in Nios IDE. Similarly, since the voice broadcast from FLASH are read out in parallel, thus a parallel to serial data conversion Verilog module is needed.

TFT LCD display: The 3.6" LCD module is the active matrix colour TFT LCD module. LTPS (Low Temperature Poly Silicon) TFT technology is used and vertical and horizontal drivers are built on the panel. Horizontal scan can be from left to right or from right to left and vertical scan can be from up to down or from Down to up. We developed the LCD Controller and displayed a Tamil text to the TFT LCD. The main CPU is the Nios II processor.

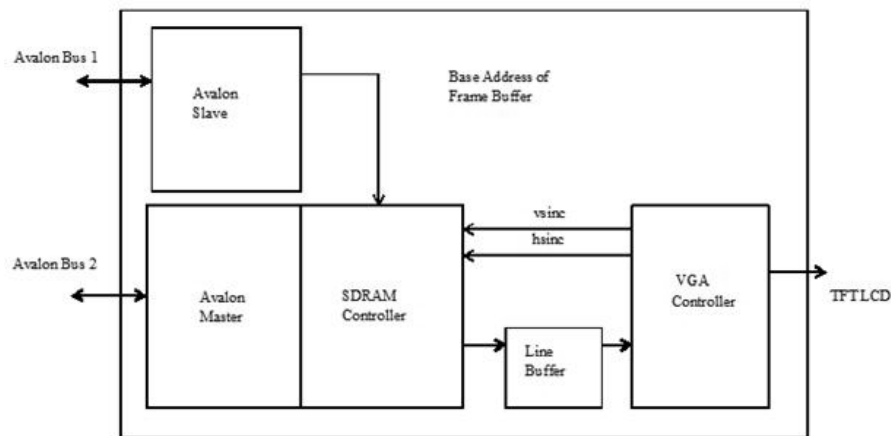


Figure. 5 TFT interfacing module

The Nios II processor connects to ext_RAM_bus, 3-D accelerator, 7-segment controller, SDRAM controller, TFT-LCD controller, and so on. The ext_RAM_bus module is a tristate Avalon bus bridge that connects the Nios II processor to flash memory and SRAM, which are the instruction memory and data memory, respectively, used to run the Nios II processor.

New Component menu option in SOPC Builder is used to include the LCD Controller module and specify its signals. The signals between the LCD Controller and Avalon bus became the pins of the Nios II module. The slave side of the module is connected to the Nios II processor and the master side is connected to the SDRAM controller.

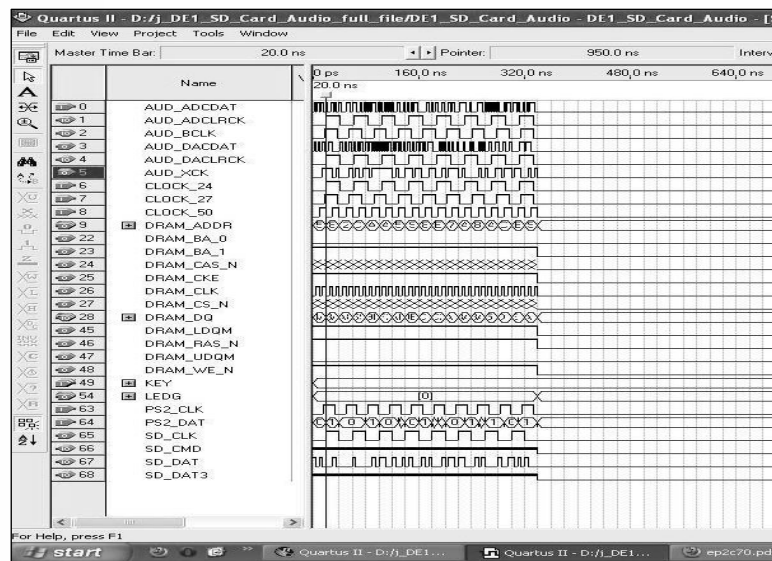
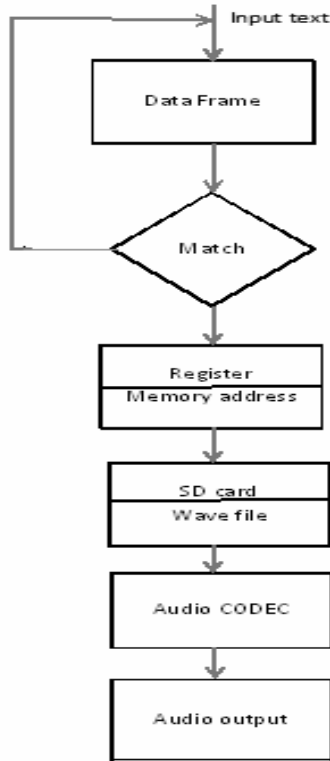
Proposed Design Flow Algorithm

- Initialize and load the acoustic library, which consists all the audio recordings;
- Load the target or input word; initialize the register counter;
- While (target letter match to memory address) do
- {

- Wave files play via audio codec from the corresponding memory address;
- when player done = 1 implies audio player has finished playing the wave file;
- /* end the loop*/
- reset the register counter ;
- repeat the procedure for all the target letter;

FIG.6 DESIGN FLOW OF THE TTS

Simulation Results



The project is simulated using Verilog in Altera tools. Simulation for the sound of 'a'(in Tamil).

Conclusion

In this paper, the design and development of functional architecture of Tamil text to speech synthesis is described. Different feature streams are experimented with the system that optimises the parameter extraction to perform efficient speech synthesis.

Future Work

This work can be extended for concatenating the Tamil syllables and also produce the natural Tamil voice from FPGA based machine.

References

1. R.Thangarajan, "Syllable Based Continuous Speech Recognition for Tamil", in South Asian language review, 2008.
2. Nageshwararao and Hema, "Text to speech synthesis using syllable like Unit", 2005 IEEE paper
3. P.Nirmala Devi and R.Asokan "VLSI implementation of speech to text conversion" in Proceedings of International conference of intelligent Knowledge systems (IKS-2004), Turkey, 2004.
4. G.L. Jayawardhana Rama and A.G Ramakrishna "A complete text to speech synthesis' system in Tamil" IEEE Workshop on Speech Synthesis, 2002.
5. B.H.Juang," Why speech synthesis?" IEEE Transactions on speech and audio processing, 2001
6. A.G.Ramakrishnan, "Thirukural text to speech synthesis system", proco. Tamil Internet 2001, Kuala Lumpur
7. Pong P.Chu, "FPGA prototyping by Verilog examples", John Wiley & Sons, 2008
8. Douglas O'Shaughnessy, "Speech Communication", Universities Press, 2004, Second edition.
9. http://www.altera.com/literature/manual/mnl_cii_starter_board_rm.pdf

FaceWaves : Tamil Text-To-Speech with Lip Synchronisation for a 2D Computer Generated Face

A.G. Tamilarasan, Madhan Karky

tamil_2206@yahoo.co.in, madhankarky@gmail.com

Department of Computer Science & Engineering

College of Engineering Guindy

Anna University

Abstract

This paper presents a system for converting Tamil text to speech and synchronizes the lip movements of a 2D computer generated face to the generated speech. This paper describes the components of the system such as, grapheme to phoneme converter, a rule-based phoneme-sound selector, wave merger, lip-phoneme synchronizer and face animator. Presenting the grapheme to phoneme algorithm and the phoneme-sound selector rules, this paper presents the results from the grapheme to phoneme converter and rule based-sound selector. The paper concludes with future extensions of the presented work.

Introduction

Tamil language has strict phonetic rules defined many thousand years ago in the ancient grammar definition Tholkaappiyam. These rules have formed the basis of many text to speech systems that exist in Tamil today. This paper proposes a Tamil text-to-speech and lip-synchronization system that takes Tamil text and a 2D computer generated face as input and converts the Text to speech waves and lip synchronize the face to speak the given text along with audio as an animated video.

The motivation behind this work is to provide such a subsystem for a pure text-to-video system where an entire video can be created purely using natural language. Such a system will be of immense use to physically challenged and also enables common people with very less computer knowledge to create and distribute animation video. One major advantage of having such a system would be to transfer such animation over Internet or mobile as a simple text and converting the text to video using a local client.

This paper is organized into four sections. The second section discusses the background and literature closely related to this work. The third section provides the system architecture and discusses the different modules and their functions. The fourth section discusses the results and issues related to text-to-speech conversion and lip synchronization. The fifth section summarizes and concludes this paper and discusses future research directions.

Background

There have been many research works related to Text-to-Speech and generally the text-to-speech systems fall in one of the two categories namely speech synthesis and phoneme concatenation. Sreekanth Majji and Ramakrishnan A. G. used a labeled database to make a system capable of producing a speech output[1]. John Lewis proposed Naive approach for automatic lip- sync to open the mouth in proportion

to the loudness of the sound [3]. Keith Waters and Thomas M. Levergood demonstrated an automatic lip-synchronization algorithm for automatically synchronizing lip motion to a speech synthesizer in real-time[5]. Masatsune Tamura, Shigekazu Kondo, Takashi Masuko, and Takao Kobayashi used parameter generation algorithm, which is used to generate an audio-visual speech parameter vector sequence[4]. Marc Schroder developed Modular Architecture(MARY) for speech synthesis and has the capability of parsing speech synthesis markup.[2]

The Text-to-Speech system proposed in this paper is based on the Tamil Voice Engine [6] proposed by Madhan Karky et al., The Tamil Voice Engine takes Tamil text as input and uses a phoneme database with over 4000 entries. The words are split into corresponding phonemes using rules and the appropriate phonemes are selected and concatenated. This paper uses a similar system for Text-to-Speech but incorporates more rules to improve the phoneme selection and thereby improves the correctness of generated speech.

System Design

The system proposed for text-to-speech and lip synchronization can be naturally split into two major components. The Text-to-speech component receives text as input and converts the graphemes to phonemes. The waves associated with the phonemes are concatenated along with pauses into a single audio wave as continuous speech. The Lip-Sync component of the system receives a 2D computer generated face with all parts of faces and their coordinates. The coordinates of lips are identified and they are modified for different lip positions for different phonemes. The system maintains a lip coordinate index for phonemes that require modifications. Figure 1 presents the overall design of the system.

Text-to-Speech: The Text-to-Speech component comprises of four modules. Text Processor, Grapheme to Phoneme Converter, Rule Based Phoneme Selector and Merger. The component also comprises of a phoneme database. These modules together convert a given text dialogue to speech.

- **Text Processor:** The Text processor processes raw input text and tags the text to appropriate tags as words and pauses. Inter-word pause, inter-sentence pause are tagged by the text processor. A document is thus converted to a set of ordered word and pause tags.
- **Grapheme to Phoneme Converter:** Grapheme to Phoneme converter takes the ordered set from Text Processor as input and processes the word tags and splits them further into phonemes. The grapheme to phoneme conversion is based on an algorithm provided in [6] that uses consonant-vowel combinations of letters in a word. An example of such a grapheme to phoneme conversion is provided below.

தமிழுக்கு அமுதென்று பேர் | thamizhukku amudhenru paer த-மி-முக்-கு அ-மு-தென்-று பேர்

Will be split into the following phoneme combinations separated by hyphen.

- **Phoneme Merger:** The phoneme database stores 4646 phonemes and based on the split provided by the grapheme to phoneme converter, Phoneme merger applies certain rules to identify the right set of phonemes for the current split.

<tha+mi+zhuk+ku> <a+mu+dhen+dru> <paer>

The selection of 'dhen' in the above example instead of 'then' is based on the rule that the sound of the phoneme changes based on where it occurs in a sentence. Similar rules are applied for the hard consonants (ka/ga sa/cha da/ta tha/dha pa/ba Ra/tra) based on their occurrences. 'kutriyalukaram' a property that states that words that end with 'u' have their meter length cut short by half a meter. This rule is also being handled as a rule for selecting appropriate phoneme. Phoneme merger then chooses the appropriate waves from the phoneme database for the phonemes selected by the Phoneme Selector. The selected phonemes are then merged along with the appropriate pauses. The phoneme merger creates a single wave file that can be attached to a video or used for any text-to-speech application. In this framework, it sends the wave to animator and the phoneme and duration information to lip synchronization modules respectively.

Lip Synchronization: The Lip Synchronization component comprises of a Face Generator, Face Animator and Lip Synchronizer. These modules together generate a video of a speaking face whose lip movements are synchronized to the speech generated by the Text-to-Speech component.

- **Face Generator:** The face generator module developed as part of the FaceWaves framework, receives textual description of facial features in Tamil. The descriptions are converted to coordinates and dimensions of various parts of the face. The coordinates of the face are used as input for the lip synchronizer
- **Lip Synchronizer:** Lip synchronizer receives temporal information about the generated wave such as the phonemes, the start and end time of every phoneme and pauses. Lip Synchronizer uses this information to suggest the lip positions for the face animation at every time instance. An example sentence and how it is converted to a phoneme-pause temporal information is given below.

Table 1: Duration of phonemes

Table 2: Lip Position

- **Face Animator:** Face Animator uses the temporal information provided by the lip synchronizer to modify the lip coordinates of the face provided by the Face Generator. Face Animator uses the speech wave provided by the Phoneme Merger and creates frames of faces with varying lip coordinates over the same timeline. As the wave timeline and animation frame timeline are one and the same, the animation video has 100% synchronization accuracy.

நான் இங்கு வந்தேன் | naan ingu vandhaen

Phoneme	From(Sec)	To(Sec)
நான் naan	0.0	0.445
Pause	0.446	0.696
இங் ing	0.697	0.931
கு ku	0.932	1.207
Pause	1.208	1.457
வந் van	1.458	1.707
தேன் thaen	1.708	2.083

Lip Position	From(Sec)	To(Sec)
2	0.0	0.445
1	0.446	0.696
3	0.697	0.931
5	0.932	1.207
1	1.208	1.457
13	1.458	1.707
6	1.708	2.083

Results

Text-to-speech system was tested over a dictionary [7] of 2,00,000 root words. Text was converted to speech with 92% accuracy. Words with non-Tamil characters and missing phonemes can be attributed to the 8% results were the right phonemes were not selected. The same system when tested with 200 named entities (person names and organization names) gives 98% accuracy.

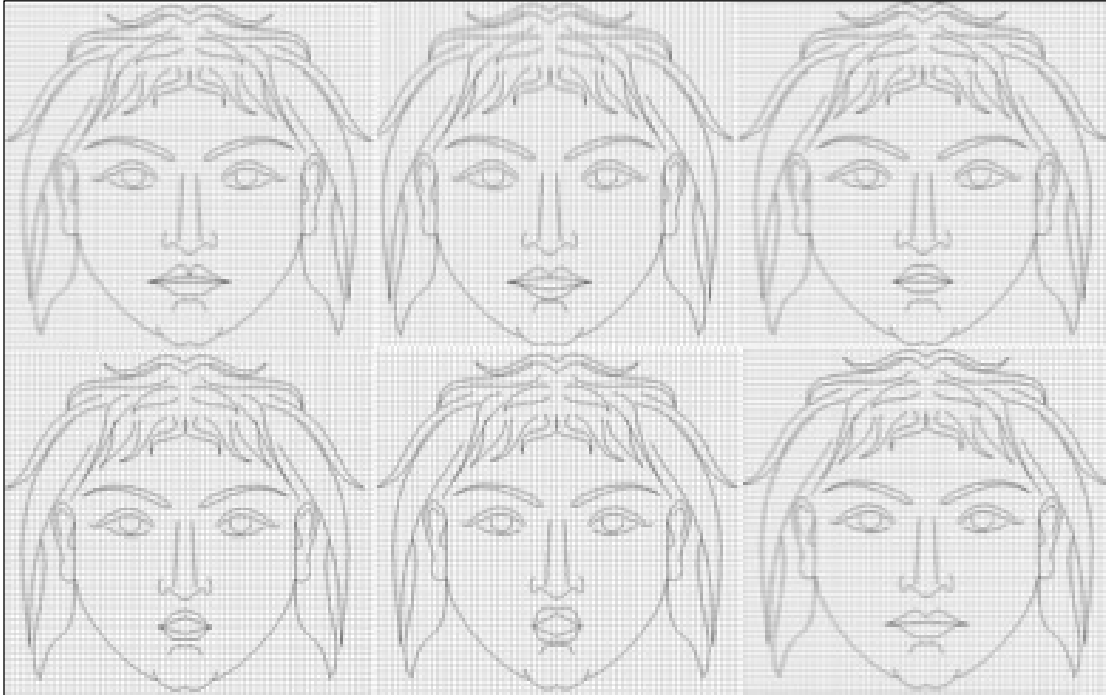


Fig 2: Lip Positions

The Lip Synchronizer and Face Animator use a phoneme to lip-coordinate mapping. The mapping provides information as lip positions. Figure 2 provides the lip positions for various phonemes. The faces in Figure 2 were generated by the Face animator with input from Lip Synchronizer. As mentioned in the previous section the lip synchronization has 100% accuracy as the same timeline is considered for the wave and for the animation.

Conclusion and Future Work

The current system has certain limitations such as the phoneme recordings are not normalized and the merged phonemes are not smoothened. Wave smoothening and professionally recording the sounds and mastering the waves will be essential to improve the quality of the audio. Secondly, the face animation is currently done for a 2D face. If the similar system can be extended to animate 3D faces the animation would look more realistic. This system we believe would be of immense use to common people with very little computer knowledge to create and distribute 2D animations. Creating mobile agents to communicate text messages or creating automatic news reporters with these animated faces will take this project to the next level.

References

1. Sreekanth Majji, *Festival Based Maiden TTS System for Tamil*, Language research paper, Indian Institute of Science, Bangalore, 2007.
2. Marc Schroder and Jurgen Trouvain, *German Text-to-Speech Synthesis System*, Institute of Phonetics, University of the Saarland, Germany, 2008.
3. John Lewis, *Automated Lip-Sync: Background and Techniques*, Computer Graphics Laboratory, New York Institute of Technology, 1991.
4. Masatsune Tamura, Shigekazu Kondo, Takashi Masuko, and Takao Kobayashi, *Text-to-audio visual speech synthesis based on parameter generation from HMM*, Interdisciplinary Graduate School of Science and Engineering Tokyo Institute of Technology, Yokohama, 2007.
5. Keith Waters and Thomas M. Levergood, *DECface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces*, Digital Equipment Corporation Cambridge Research Lab, 1993.
6. Madhan Karky V, Sudarsanan N, Thiyagarajan R, Manoj Annadurai, Dr.Ranjani Parthasarathi and Dr.T.V.Geetha: *Tamil Voice Engine*, Tamil Internet Conference, Malaysia, 2001.
7. Agaraadhi Online Tamil Dictionary, www.agaraadhi.com, last accessed 20/04/2010.

Classical Encryption Techniques for Tamil Text

P.Navaneethan Dept. of Electrical & Electronics Engineering PSG College of Technology Coimbatore - 641004, India. pnn@eee.psgtech.ac.in	C.L.Brindha Devi Dept. of Computer Applications K S Rangasamy College of Arts & Science Tiruchengode-637209, India. clbrindhadevi@gmail.com	N. Karthikeyan IInd MCA, Dept. of Maths and Computer Applications PSG College of Technology, Coimbatore-641004 karthikeyan513@gmail.com
---	---	---

Abstract

This paper describes the security aspects of the Tamil text that is being stored and transmitted over the Internet. The character set in Tamil language shall be categorized into frequently used Tamil characters and infrequently used Tamil characters. The frequently used Tamil characters are divided into consonants, vowels and combined characters. This paper describes how to encrypt the frequently used Tamil characters using a 16-bit Crypto Index. The Crypto Index serves dual purpose, namely, one to find the algorithm to be used and the other to specify the key to encrypt the consonants & vowels.

Keywords: Encryption, Decryption, Substitution, Crypto Index, Rotation, and Mirroring

Introduction

A novel approach for the encryption and decryption of Tamil text is proposed. A 16-bit Crypto Index is used to select the encryption technique and the corresponding key to be used. Though, an arbitrary substitution of Tamil character with yet another set out of 247! ways is feasible, remembering such a substitution based key is impossible, though equally impossible would be the breaking of one such ciphered text. In Tamil text one will not find things like digram, trigram etc. So it eliminates the brute-force technique for cryptanalysis of ciphered Tamil text. In general, when the size of the key is large, it needs to be stored in a media, which makes the whole scheme insecure. The Crypto Index scheme is sufficiently complex, and the key can be kept confidential to oneself.

CV Based Encryption

In this scheme, only those basic characters, namely, consonants and vowels, which are used to derive all the phonetic characters, are encrypted. For example, the Tamil word, திருமணம் can be represented in CV form as த்இர்உம்அண்அம். A substitution like த் -> க், ற் -> ம், ம் -> ன், ண் -> ஞ் for the consonants and அ -> ஓள, இ -> ஊ, உ -> எ would have encrypted the CV representation into க்ஊம்என்ஓளஞ்ஓளன் and hence, திருமணம் will be ciphered as கூமெனெளஞெளன், which means that the ultimate substitution within Tamil alphabet set as தி -> கூ, ரு -> மெ, ம -> எனெ, ண -> ஞெள, ம் -> ன்.

Substitution Schemes

a. Rotation Based Substitution [R]

The given text is converted into CV form. Each character in the CV form is rotated 'k' places further down the characters in its set, where 'k' does the role of a key. For example, consider the vowel character set with the key as 4. Encryption gives the following result.

Position	:	0	1	2	3	4	5	6	7	8	9	10	11
Plain Alphabet	:	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
Ciphered Alphabet	:	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	அ	ஆ	இ	ஈ

The mathematical model for this encryption is:

$$\lambda(x) = (x + k) \bmod n$$

where,

x -- position of the character to be encrypted.

n -- total number of characters in the vowel or consonant character set.

k -- takes the value in the range 1 to 11 for vowels and 1 to 17 for consonants.

Decryption gives the following result.

Position	:	0	1	2	3	4	5	6	7	8	9	10	11
Ciphered Alphabet	:	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
Plain Alphabet	:	ஐ	ஒ	ஓ	ஔ	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ

The mathematical model for decryption is:

$$\lambda^{-1}(x) = (x - k) \bmod n$$

b. Mirroring Based Substitution [M]

The mirroring can be either with respect to a single axis or multiple axes. In Single Axis Mirroring technique, the given character set of size 'n' is divided into two equal halves by an axis. The character at position 'n-1' is replaced by character at position '0' and 'n-2' by '1' and 'n-3' by '2' etc. For example, consider the vowels set where n = 12.

		<u>j=0</u>					<u>j=1</u>						
Position	:	0	1	2	3	4	5	6	7	8	9	10	11
Plain Alphabet	:	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
Ciphered Alphabet	:	ஔ	ஓ	ஓ	ஐ	ஏ	எ	ஊ	உ	ஈ	இ	ஆ	அ

The mathematical model for Multiple Axes Encryption is: $\gamma(i) = (2j + 1)(n/m) - 1 - i$,

where, i lies in the closed interval $[j(n/m), (j+1)(n/m)-1]$, n is the number of characters in the character set, m is the number of axes used for mirroring, for $j = 0, 1, \dots, (m-1)$.

c. Transposition Based Substitution

In this scheme, the given character set is stored row-wise in an $(r \times m)$ matrix, where $n = r \times m$. The key δ is a reordering of $[0, 1, \dots, m-1]$, and is used to encrypt the given character set by assigning to each column the corresponding mapping. Each column of the matrix is then read based on the ascending order of the key values which results in the reordering of the character set. The 18 Tamil consonants are arranged as a (6×3) matrix as in Fig.3.1. In case, δ happens to be $[1, 2, 0]$ then, the resulting reordering permutation α is obtained by reading each column based on the ascending order of the key δ ; i.e., as per $\delta^{-1} = [2, 0, 1]$. The numerical value in each cell of the matrix specifies the position of the consonant in the character set, as shown in Fig.3.1.

Position : 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
 Plain Alphabet : க் ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ண்

	1	2	0
க்	0	ங்	1
ச்	2	ச்	3
ஞ்	3	ட்	4
ண்	4	ண்	5
த்	6	ந்	7
ப்	8	ப்	9
ம்	9	ய்	10
ர்	11	ர்	12
ல்	12	வ்	13
வ்	13	ழ்	14
ழ்	14	ழ்	15
ள்	15	ற்	16
ற்	16	ண்	17

Fig. 3.1 Row-Wise Distribution of Consonants

$\alpha_{\chi} =$
 க் ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ண்
 ச் ண் ப் ர் ழ் ள் க் ஞ் த் ம் ல் ள் ங் ட் ந் ய் வ் ற்

The above permutation can be grouped into 'm' groups; i.e., in this case as G0, G1, G2.

Here, $\text{floor}(i/r)$ gives the group id 'x'; i.e., $x \in \{0, 1, 2, \dots, m-1\}$, and hence $\delta^{-1}(x)$ gives the starting value for the reordering associated with the group x . Moreover, $(i \% r)$ gives a measure of as to how far 'i' is from the start of the group. For example, 14 is 2 units away from the start, namely 12, in G2; this is nothing but $14 \% 6$. Thus, herefore $\delta^{-1}[\text{floor}(i/r)] + [i \% r]m$ gives the reordering value for 'i' in group x , and the model for encryption is : $\alpha(i) = \delta^{-1}[\text{floor}(i/r)] + [i \% r]m$,

where $r = n/m$ is the number of rows, n total number of elements, m number of columns

$\alpha(14) = \delta^{-1}[\text{floor}(14/6)] + [14 \% 6]*3$; i.e., $\alpha(14) = \delta^{-1}(2) + 2*3 = 7$; i.e., $\alpha(\text{ழ்}) = \text{ந்}$.

In order to get α^{-1} , the given character set is stored in an (r x m) matrix column-wise, as shown in Fig. 3.2. Then, the elements in δ^{-1} are assigned to the columns of the matrix. Each row of the matrix is then read based on the ascending order of the key⁻¹ value; i.e., δ^{-1} , row-wise. The numerical value in each cell of the matrix specifies the position of the consonant in the character set.

	2	0	1
க	0	த்	6
ங்	1	ந்	7
ச்	2	ப்	8
ஞ்	3	ம்	9
ட்	4	ய்	10
ண்	5	ர்	11
			ல் 12
			வ் 13
			ழ் 14
			ள் 15
			ற் 16
			ன் 17

Fig. 3.2 Column-Wise Distribution of Consonants

$$\alpha_{\chi}^{-1} = \begin{matrix} \text{க ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ண்} \\ \text{த் ல் க் ந் வ் ங் ப் ழ் ச் ம் ள் ஞ் ய் ற் ட் ர் ண் ண்} \end{matrix}$$

The mathematical model for decryption is $\alpha^{-1}(i) = \delta [i \%m]r + [\text{floor}(i/m)]$, where $r = n/m$ is the no. of rows, n \diamond total no. of elements, m \diamond no. of columns. For the transposition-substituted value 10, its inverse is 15. This is obtained by applying the decryption model; i.e., $\alpha^{-1}(10) = \delta [10\%3] 6 + [\text{floor}(10/3)] = \delta (1) 6 + 3 = 2*6 + 3 = 15$ i.e., $\alpha^{-1}(y) = ள்$.

Similarly, it can be applied to vowels also.

d. Concatenation of Substitutions

As each substitution (one vowel with another or one consonant with another) could be represented as a permutation, these substitutions can be concatenated so as to arrive at another substitution; i.e., Let α stand for transposition based substitution and β stand for rotation based substitution. Then, $\alpha \circ \beta$ (read as α composition β) stands for a new substitution γ , where rotation is applied first, followed by transposition. Mathematically, if $\gamma = \alpha \circ \beta$, then, $\gamma(i) = \alpha(\beta(i))$. For example, let α represent transposition based substitution, where δ is [2,0,1] and β represent the set of consonants undergoing left shift by 2 positions and vowels by 9 positions.

$$\alpha_{\chi} = \begin{matrix} \text{க ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ண்} \\ \text{ங் ட் ந் ய் வ் ற் ச் ண் ப் ர் ழ் ண் க் ஞ் த் ம் ல் ள்} \end{matrix}$$

$$\beta_{\chi} = \begin{matrix} \text{க் ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ன்} \\ \text{ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ன் க் ங்} \end{matrix}$$

$$\alpha_{\theta} = \begin{matrix} \text{அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ} \\ \text{ஆ உ ஏ ஒ இ ஊ ஐ ஔ அ ஈ எ ஒ} \end{matrix}$$

$$\alpha_{\theta} = \begin{matrix} \text{அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ} \\ \text{ஆ உ ஏ ஒ இ ஊ ஐ ஔ அ ஈ எ ஒ} \end{matrix}$$

$$\beta_{\theta} = \begin{matrix} \text{அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ} \\ \text{ஒ ஓ ஔ அ ஆ இ ஈ உ ஊ எ ஏ ஐ} \end{matrix}$$

Let, $\gamma_{\chi} = \alpha_{\chi} \circ \beta_{\chi}$ and $\gamma_{\theta} = \alpha_{\theta} \circ \beta_{\theta}$. Then,

$$\gamma_{\chi} = \begin{matrix} \text{க் ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய் ர் ல் வ் ழ் ள் ற் ன்} \\ \text{ந் ய் வ் ற் ச் ண் ப் ர் ழ் ன் க் ஞ் த் ம் ல் ள் ங் ட்} \end{matrix}$$

$$\gamma_{\nu} = \begin{matrix} \text{அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ} \\ \text{ஈ எ ஒ ஆ உ ஏ ஒ இ ஊ ஐ ஔ அ} \end{matrix}$$

e. Crypto Index Scheme

To encrypt the consonants and vowels, a novel scheme called Crypto Index Scheme is proposed. In this scheme, a 16-bit key is used , which provides indices to the encryption techniques to be employed. The key is grouped into four groups, namely, G1, G2, G3, G4. The bits b0, b1, b2, b3 (G1) are used to encrypt the vowels based on a Rotation [R], the next 5 bits, namely, b4, b5, b6, b7 and b8 (G2) are used to encrypt the consonants based on a Rotation, and the next 6 bits, namely, b9, b10, b11, b12, b13, b14 (G3) are used to go for further concatenation with either Transposition based substitution or Mirroring [M] based substitution, which is applicable to Consonants(C) and/or Vowels (V) based on its value. G4 is used to

specify the type of rotation (left/right).For example, if the key is 0 000111 00001 0001 then, it can be divided into four groups as follows:

0 000111 00001 0001
G4 (b15) G3 (b14-b9) G2 (b8-b4) G1 (b3-b0)

G1 -- is used to encrypt only Vowels, G2 -- is used to encrypt only Consonants

G3 -- is used for further concatenation of substitutions applicable to Consonants and /or Vowels

G4 --Type of rotation (Left or Right)

G3 is converted into a radix-4 number, say F. Based on the radix-4 digits, an appropriate algorithm is chosen for encrypting the Vowels and/or Consonants. The radix-4 digits be symbolically denoted as F1, F2 ,F3. If F1F2F3 is,

123 \diamond 0 missing \diamond Mirroring on Consonants &Vowels,

023 \diamond 1 missing \diamond Transposition of Vowels, 013 \diamond 2 missing \diamond Transposition of Consonants

012 \diamond 3 missing \diamond Transposition of both Consonants & Vowels

After choosing the algorithm, F_i should contain only radix-4 values. Since it is to be used as a key (δ) for transposition, replace the occurrence of 3 by 2 in F. If G3 has the combination of two 0's and 1/2/3 then, the algorithm is applied to both the consonants and the vowels. The consonants and vowels are divided into 3 groups and for the group having the F_i value 1/2/3, mirroring is applied, and for the other 2 groups rotation by 1/2/3 times is applied based on G4. If G3 has 000 then no substitutions are made.

If G3 has the combination of two 1's and 0/2/3, then the algorithm is applied only to vowels. The character set is divided into 3 groups and if G3 is 111 then each group is rotated 1 time and for the group having the F_i value 0/2/3, mirroring is applied, and for the other 2 groups rotation by 0/2/3 times is applied based on G4. Here each group is split into 2 parts and while splitting, the 1st part length should be equal to one more than the value 0/2/3.

If G3 has the combination of two 2's and 0/1/3, then the algorithm is applied only to consonants. If G3 is 2 2 2 then a three 6-Consonants group is formed. Each group is rotated 1 time and for the group having the F_i value 0/1/3, mirroring is applied, and for the other 2 groups rotation by 0/1/3 times is applied based on G4. Here each group is split into 2 parts and while splitting the 1st part length should be equal to one more than the value 0/1/3.

If G3 has the combination of two 3's and 0/1/2, then the algorithm is applied to both consonants and vowels. If G3 is 3 3 3 then rotation applied to both Consonants & Vowels group. Each group is rotated 1 time. Here the group having the F_i value 0/1/2, mirroring is applied, and for the other 2 groups rotation by 0/1/2 times is applied based on G4.

So the Crypto Index Scheme defines 64 (24+40 = 26) different crypto indices for the group G3.

தமிழ் இணையம் is encrypted using Crypto Index Scheme as follows. தமிழ் இணையம் is first Converted to CV Form as த்அம்இழ் இணையம்அம். If the key is 0 100001 00010 1001 then

G1 = 1001, G2 = 00010, G3 = 100001, G4 = 0. After applying rotation followed by transposition, we get
□□□i.e., □c on Consonants & □v on Vowels. Hence,

த அம் இழ் □ ப் ஈன் ஒல் □ பீனொல் □ இண் ஐய் அம் □ ஒண் ஊக் ஈன் □ ஒணூகீன்

The ciphered text for தமிழ் இணையம் is பீனொல் ஒணூகீன்.

Conclusion

This paper has proposed and illustrated a novel scheme for encrypting the Tamil text using Crypto Index Scheme. The frequently used Tamil character set has 247 characters and hence, 247! possible ways of substitutions, where the key (i.e., the arbitrary substitution) cannot be remembered, in general. Hence, methods have been derived to identify substitution schemes based on a 16-bit Crypto Index, the key that can easily be remembered, but, which makes the schemes complex enough. Moreover, Tamil language has no digrams, trigrams etc. as English has. Hence, it eliminates the brute-force attack by a cryptanalyst.

References

1. Stallings, Cryptography and Network Security, Pearson Education ,Third Edition, 2003.
2. Behrouz A. Forouzan, Cryptography and Network Security, Tata McGraw-Hill, 2007
3. Navaneethan P., Madheswaran R., Balasubramanianm R., and Bharathidasan R.V., "PANDITHAM: An Optimal Character-oriented Protocol for Multilingual Computing", Tamil Inayam- 2000, Singapore.

Statistical Analysis and Visualization of Tamil Usage in Live Text Streams

J. Jai Hari Raju, P. IndhuReka, Dr. Madhan Karky

jaihari1989@gmail.com, p.indhu@gmail.com, madhankarky@gmail.com

Department of Computer Science & Engineering

College of Engineering, Anna University, Guindy

Abstract

Tamil is slowly gaining popularity as an active language in social networking on the Internet. This paper aims at statistically analyzing the usage of Tamil words in text streaming social networking sites. This paper proposes an active Tamil text stream reader designed to obtain live usage statistics of Tamil words in Twitter, a text based social networking service. A spatio-temporal dynamic index is maintained by the text stream reader and the word usage and the geo-tags are recorded along with a time stamp. The paper also presents a visualization tool where the data captured in the spatio-temporal dynamic index can be visualized graphically to show what topics are gaining and losing popularity over time and space. The results from the analysis are discussed along with snapshots from the visualization tool. The paper concludes with open questions for future research in active text stream analysis for Tamil language.

Introduction

The number of text streaming social networking sites has increased drastically and the usage of Tamil in these sites has reached considerable proportions. The analysis of which will reflect the current scenario of Tamil usage among web users. These text streams can be analyzed to obtain the usage statistics of words over time and geographic location. This analysis requires a tool which will collect data pertaining to a given Tamil word, analyze it then index it. This paper proposes an overall architecture of such a framework.

The proposed framework contains a text stream retriever which interacts with the site in order to obtain the text stream units for analysis. The obtained data is parsed based on various constraints and the information retrieved is stored in an index. This information is then subjected to analysis.

The results from such an analysis reflect various aspects of Tamil's web usage. When the results are visualized with respect to time, they depict how the usage frequencies of words vary. This data can also be used to find which topics are gaining and losing importance. When the results are analyzed with respect to geographic location they provide insights into the global scenario of Tamil. In section 2 we provide an overview of the literature survey conducted. In section 3 we discuss the design of the various modules of the framework which provides us an overview of the spatio temporal properties of Tamil text streams. In section 4 we discuss the implementation of the proposed framework and the results obtained from the analysis. Finally we conclude in section 5 with directions for further studies in the field of text stream analysis for Tamil language.

Background

In the literature there are existing works on text stream analysis. Qiankun Zhao et al. explored the temporal and social information of text streams in order to achieve better results in event detection [1]. Jon Kleinberg analyzed the various techniques for Topic Detection and Tracking and also the various information visualization techniques for interpreting data from a temporal dimension [2]. Nilesh Bansal et al. developed a system called BlogScope [3]. BlogScope is an information discovery and text stream analysis system which follows blogs and analyses their content. Danyel Fisher et al. developed a visualization to track narrative events as they develop in text streams [4]. Le Wang et al. proposed a double time window algorithm for conversation extraction in dynamic text message streams [5]. Vagelis Hristidis et al. explored the techniques for extracting useful information from a collection of text streams and proposed a system for keyword search on textual streams [6].

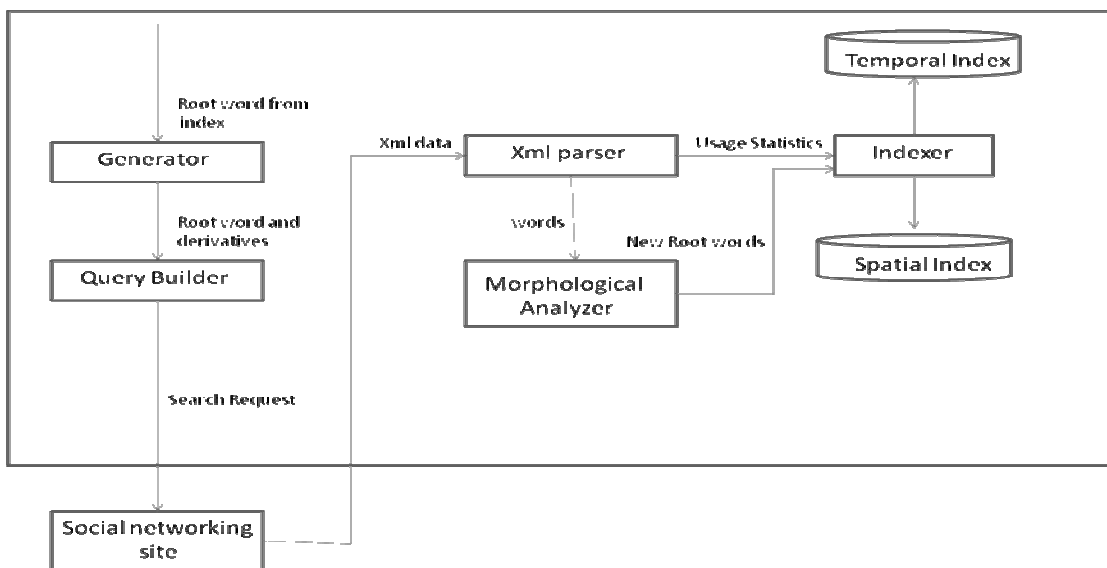
Design of Tamil Text Stream Analyzer

The Tamil Text Stream Analyzer consists of the following major components:

- Text Stream Retriever
- Analyzer
- Indexer

Figure (1) given below depicts the architecture of the framework.

Figure 1: Tamil Text Stream Analysis Framework



a. Text Stream Retriever

This module is responsible for retrieving the text stream usage instances for a given word. The initial input for this module is obtained from a list of predefined popular root words. In order to prevent the overhead of storing a very large search list comprising of both the root words and their derivatives only the root words are stored and their corresponding derivatives are obtained using a generator. The search query is constructed using the OR operator so that the occurrence of even one of the forms is taken in to account. Storing of only the root words in order to increase the memory efficiency is an important design decision. The query is targeted towards the particular site's Search API. The search results returned by the APIs which are generally XML files are in turn handed over to the analyzer module.

b. Analyzer

The analyzer module is responsible for analyzing the raw data retrieved by the reader. The analyzer parses the given XML data in order to extract the required statistics. Predefined tags containing the required information are identified. The total number of entries for the given word is counted. As the amount of usage history maintained by the Social Networking sites is found to vary for each word, this parameter will not suffice for a conclusion to be drawn. So to provide a uniform interpretation, this module also analyses the per day usage for a fixed number of days in the past for all words. It is also possible to record the origin of the text stream unit provided the sites support geotagging. Currently Twitter, a Social networking site supports geotagging by allowing queries based on the origin of a text stream unit.

c. Indexer

As the usage statistics of words are determined for two parameters namely time and geographic location, spatiotemporal indices are used to store the data. The decision to store, search and update a fixed list of words would result in a static application which cannot be scaled. Hence a dynamic approach is chosen, wherein the text stream instances returned for a keyword are parsed to obtain new words whose root words do not exist in the search list currently. Those words are added to the search list and are analyzed in the future searches.

Index Structure: As text streams are analyzed in a spatio temporal manner, two indices are maintained. The temporal index stores the usage statistics of words with respect to time. The spatial index stores the usage statistics of words with respect to geographic location. A Hashtable data structure is used for the indices. The temporal index is built with the temporal component (Date) as the key.

Implementation

In order to implement the proposed framework, Twitter a prominent Social Networking Site is chosen as the source for text stream. The root word to be searched is taken from the index and the word's derivatives are found using a generator. The search query for the temporal analysis is constructed from those words and is posted to the Twitter Search API. In order to analyze the word usage spatially we count the number of times the word originates inside TamilNadu. The search API responds with XML data. The analyzer module parses this data and populates the statistical and temporal indices maintained. The text stream units are parsed and the words are subjected to morphological analysis and the resulting root words if not already present, are added to the index

a. Results

As the source of the text streams is a social networking site (Twitter), they contain conversations which have many stop words. These stop words are not excluded from the search. The search for stop words yields many other words which accompany them in conversations. The analysis also shows that a considerable proportion of the text streams originate outside Tamil Nadu. The user interface is designed with options for the users to visualize the results in various aspects. Options for viewing the results for the top 5 words of the week, Comparing the usage statistics of the given words and analyzing the spatial usage distribution of a given word are provided. Figure (2) given below is a screen shot of the text stream analysis system. The bar graph depicts the usage statistics of the top 5 words for a time span of 7 days from 18th of April 2010. The line graph below compares the occurrence count for a given word originating within TamilNadu against the total occurrence count.

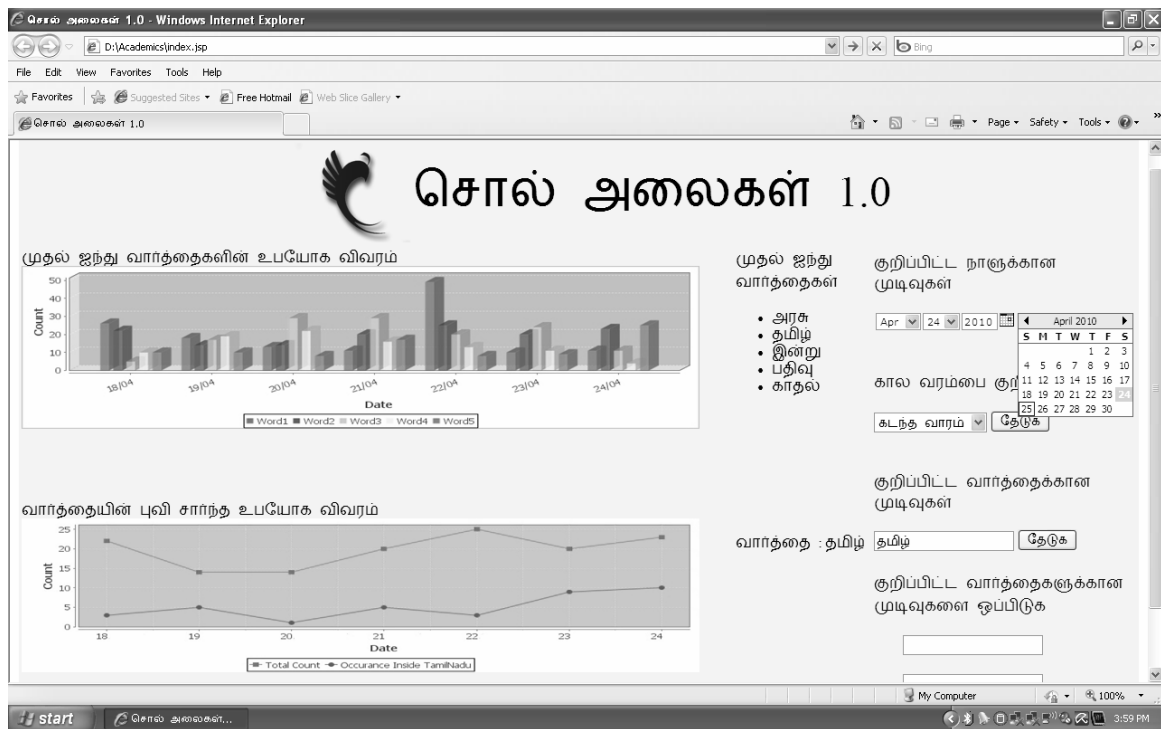


Figure 2: Screenshot of the Text Stream Analysis System

Conclusion and future work

In this paper we have proposed a framework for spatiotemporal analysis of Tamil text streams. The performance of the current system can be improved by adapting a more efficient indexing mechanism. As a next level of analysis, Topic Detection and Tracking (TDT) and prediction of evolving topics can be applied for Tamil text streams. The usage of Tamil in blogs and search engine queries is on the rise. Blogs and search engine queries have inherent temporal properties, hence this analysis can be extended to these areas also. Related work has been carried out for text streams in English. But Tamil being a highly inflectional language, requires customized searching and indexing mechanisms for efficient analysis.

References

1. Q. Zhao, P. Mitra and B.Chen, "Temporal and Information Flow Based Event Detection from Social Text Streams," 22nd AAAI Conference on Artificial Intelligence, (AAAI'07), Vancouver, Canada.
2. J. Kleinberg, "Temporal Dynamics of On-Line Information Streams," In Data Stream Management: Processing High-Speed Data Streams, (M. Garofalakis, J. Gehrke, R. Rastogi, eds.), Springer, 2004.
3. N. Bansal and N. Koudas, "Blogsphere: A system for online analysis of high volume text streams," In VLDB, pages 1410-1413, 2007.
4. D. Fisher, A.Hoff, G.Robertson and M. Hurst, "Narratives: A Visualization to track Narrative Events as they develop," In IEEE Symposium on Visual Analytics and Technology(VAST 2007).
5. L. Wang, Y. Jia and Y. Chen.(2008,Oct). Conversation extraction in dynamic text message stream. Journal of Computers (JCP). 3(10).
6. V. Hristidis, O. Valdivia, M. Vlachos, P.S. Yu, "A system for keyword search on textual streams," Proceedings of the Seventh SIAM International Conference on Data Mining, Minnesota, USA, April 26-28,2007.

An Analysis of Various Types of Distortions of Tamil Scripts

R. Indra Gandhi
Research Scholar
Dept. of Computer Science
Mother Teresa Women's
University, Tamil Nadu, India.
shambhavi.rajesh@gmail.com

Dr. K. Iyakutti
CSIR Emeritus Scientist
School of Physics
Madurai Kamaraj University,
Tamil Nadu, India.
iyakutti@gmail.com

Dr. C. Jothi Venkateswaran
Head of the Dept
Department of Comp. Science,
Presidency College, Chennai,
Tamil Nadu, India.
jothivenkateswaran@yahoo.co.in

Abstract

While reading old documents, it is difficult to read the content if the pages or print found to be in poor condition. The damage might be because of poor maintenance, poor paper quality, prolonged disturbances like blurring of ink and damages done by bookworms like silverfish and booklouse. The objective of this study is to ascertain the distortions on Tamil scripts by various sources. The factors that cause each kind of distortion, the problems associated with them and the possible solutions for the identification of each kind of distorted text have been discussed in detail. This is extremely useful for researchers engaged in recognizing the distorted documents in any script as same kind of distortion can be found in most of the scripts used globally.

Introduction

The progress of any OCR recognition, which registers the efforts of researches of the last six decades, can lead to the accuracy of 99.90% obtained with the help of any commercial OCR system when the document images are sharp, clear and noiseless. Still, there are several applications where the recognition process miserably fails when a poor image source is used. In addition, even a slight distortion of the image quality can make the accuracy of document recognition fall flat. The occurrences of distortions in scanned images are affected by various factors, which are categorized into four areas [1, 2]. Distorted documents do not include all the ideal properties of a document. Even a slight deter in the quality of source documents results in a downfall in the entire recognition process. Some well-known causes for deterioration on document images include: -

- (a) Natural calamities
- (b) Vertical cuts caused paper folding
- (c) Usage of poor quality of ink
- (d) Excessive dusty noise
- (e) Large ink-blobs merging the disjoint characters or components
- (f) Disconnection of arbitrary direction due to paper quality or the presence of foreign material
- (g) Floating of ink to the opposite pages or the next pages etc.,

(h) Defects caused during printing

(i) Defects introduced during digitization through camera, copying through photocopier and fax machines

General ink-blobs are nothing but the random occurrence of black or white pixels at or about a coordinate point. Even the spread of normal distribution (essentially the standard deviation) is also variable. Disconnection of characters is caused by the white blobs whereas black blobs results in merged characters. Keeping above in mind different kinds of distortions have been observed in Tamil document scripts.

Review Of Literature Over Different Type Of Distortion

In order to reach the goal, an ample study of research outcome in several related areas were surveyed.

Touching characters: Segmentation is a major problem during recognizing touching characters. Utilizing projection profiles and topographic method features extracted by Lee et al [3] have dealt with segmenting the touching characters. Bose and Kuo [4] used a robust structural analysis technique. Tsujimoto and Asada's [5] constructed segmentation method by several candidates for break positions. Casey and Nagy [6] utilized recursive procedure to decompose all blocks wider than a certain adaptive threshold. Hong [7] endorse the focus on segmentation of Roman script characters. Kahan et al [8] made an attempt on double differential function. A wide-ranging study of research outcomes in touching characters in Indian scripts is seen in [9-16].

Broken: Whichello and Yan [17] introduced a reconstruct method. Bern and Goldberg [18] advocates a scanning process adopting a probabilistic model. Akiyama et al [19] has taken their methods absorbed multi resolutions pyramid and fuzzy edge detectors. Lu et al. [20] proposed an algorithm based on estimation procedure and a sequential merging procedure. Nakamura et al [21] and Okamoto et al [22] in-took the propagation and shrinking in vertical and horizontal directions. The estimation of the pitch and location of pitch window helped Yanikoglu [23]. To rejoin the appropriate connected components Droettboom [24] used a technique based on graph combinatory.

Heavily Printed: Even advanced OCR of Roman script stands a failure to recognise the heavily printed characters [25].

Double-sided: Leedham et al [26] attempted the recognition process with the introduction of binarization methods with bleed through defects. Anna Tonazzini et al [27] have drawn more general approaches and statistical methods such as Independent Component Analysis (ICA) and Bline Source Operation (BSS). Dubois and Pathak [28] real samples are used for various distortion models. Like [28] more information regarding this can be accessed using [29]. Gang Zi and Doermann [30] and Gang Zi [31] proposed types of defects taking the base of blurring and mixing techniques.

Faxed document: Bloomberg [32] delivered his message as storing the fax as an electronic image. Oguro et al [33] gave a proposal of three-step solution for restoring fax documents. Randolph and Smith [34] comprised of directional components developed a binary Directional Filter Bank (DFB). The recognition accuracy has gained better percentage by the hands of Hobby and Ho [35]. Natarajan et al [36] have trained the system on distorted documents and then adapted to adjust the parameters of the trained model.

Type written document: Cannon et al [37, 38] have advanced an automatic quality improvement technique for recognition of distorted typewritten images. A new cost function was formulated by Rodriguez et al [39] to segment degraded typewritten digits.

Different kinds of distortion in Tamil scripts

A careful scanning of 200 Tamil documents resulted in the following discussion on the causes of each kind of distortion, the problems associated with them and the immediate need of analyzing the solution to overcome those problems that are obstacles of character recognition. The following enlists the various kinds of distortion.

1. Merging Characters (Touching)
2. Fragmented Characters (Broken)
3. Over Imprinted Characters (Heavily Printed)
4. Electronically Shared documents (Fax etc.,)
5. Double-Sided Documents (Bleed Through)
6. Type Written Documents

1. Touching Characters

This is the most commonly found distortion in hand-written and printed Tamil scripts. It happens due to overlapping of parts of two characters in one or more places in different zones. Segmentation is a process involved in recognition. As OCR depends heavily on the accuracy of the segmentation process, each OCR system has to perform well to maximize the result of segmentation. Documents containing touching characters are magazines with heavy printing, hand written documents and Photostatted documents copied on low quality machines.

A careful analysis and investigation help to classify the touching characters into three types.

- (a) Single touching
- (b) Multiple touching
- (c) Long touching

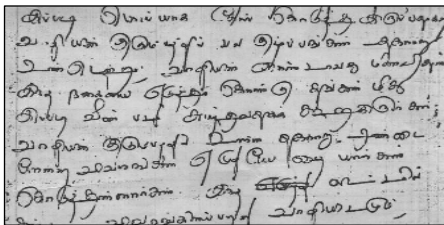


Figure 1: Touching Characters in Tamil Scripts

In the first case, the neighbouring characters touch each other in only one place, in the second case the merging occurs in more than one place and in the third, the characters merge with each other as hardly separable components. The OCR meets with a drastic decrease in accuracy when touching characters are involved.

A statistical analysis enlists the following observations

- (a) Probability is more for the merging of characters at the middle zone rather than upper or lower zones.
- (b) In most cases, touching characters in Tamil script either closely resemble some other character or totally differ from valid characters.
- (c) Most of the images encounter a single black run at the touching position.
- (d) Image characters, which are taken for treatment, are generally of two characters. Merging with one another and overlapping of three or more characters takes rare occasions.
- (e) The aspect ratio (width divided by height) clearly distinguishes the touching characters from the usual and isolated characters, which are comparatively smaller than that of the merged ones.
- (f) In fact, the identification of vertical thick ink blob differentiates the touching characters that possess the abnormal thickness in it.
- (g) From a series of analysis of touching characters, it has been noted that generally a word may contain only a pair of characters touching each other.
- (h) Possibility of more than two characters touching is very feeble.
- (i) Obviously, the characters of many Indian scripts contain a sidebar at the right end, which is absent in Tamil scripts. Hence, ambiguity is comparatively the least in that particular position.

2. Broken Characters

Resembling the touching characters, the broken characters also cause some obstruction in the character recognition of Tamil scripts.

The following observations have been made on the statistical analysis of broken Tamil characters.

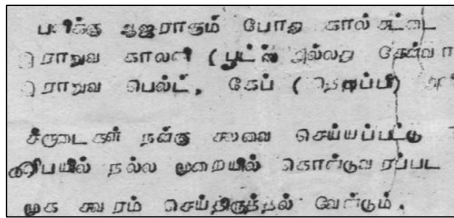


Figure 2: Broken Characters in Tamil Script

- (a) Broken characters have common occurrence more in lower zones rather than upper and middle zones.
- (b) The appearance of a character is generally not similar in shape like other individual characters. In some cases it may be same in shape of some other character.

- (c) Despite the distinct shapes of each character, some resemble other, causing complexity of conception.
- (d) Generally, there is a heavy loss of information in Tamil when compared with headline-based characters.
- (e) Broken characters generally adhere to the aspect ratio (width divided by height) lesser than other single isolated characters.
- (f) The segmentation of vertical and horizontal break of the characters has added one more problem of diagonally broken characters.
- (g) The improper spacing of character causes overlapping lines producing difficulty in understanding.

3. Heavily Printed

Characters

Sometimes isolated characters turn out to be unidentifiable due to heavy printing, equally significant problem like touching characters. So the problem of heavily printed character of Indian language still poses a need for an innovative solution. This kind of distortion takes the same sources as the touching characters.

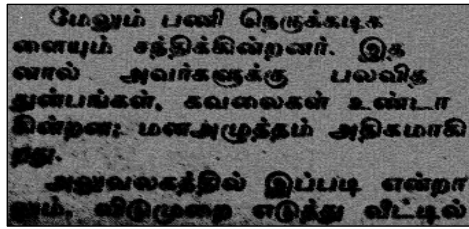


Figure 3: Heavily Printed Characters in Tamil Script

Observations after statistical analyses are enlisted below:

- (a) Generally, the shape of a heavily printed character may look like some other character.
- (b) In most of the characters, when printed heavily, there is a gain of a loop in their structure.
- (c) Heavy prints cause damages in almost all zones. Even in clean documents heavy printing sometimes heavily influence the character recognition.
- (d) Extending the problem, heavy printing leads to merging of characters, which fall into the touching character category.
- (e) Unfortunately the heavily printed characters also occupy the same space as the original characters and hence maintain the same aspect ratio.
- (f) Practically, it is difficult to extract the features of heavily printed characters due to their likeness with the originals. They take just a blob of the pixels with height and width of the original characters with no ascenders or descenders to help distinguish them.
- (g) The reasons of production of heavily printed characters are as same as that of touching characters.

4. Double-Sided Documents

This is a kind of distortion found generally in very old documents where a text on a side is visible on the other side, which is technically called as show-through or bleeds-through problem. This is one of the most challenging one in distortion problems.

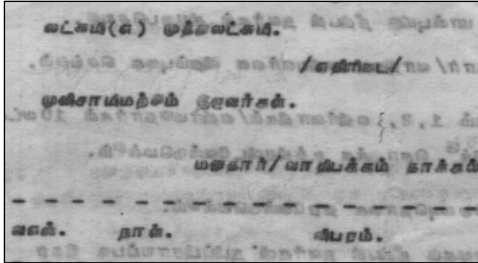


Figure 4: Backside Text Visible In Tamil Script

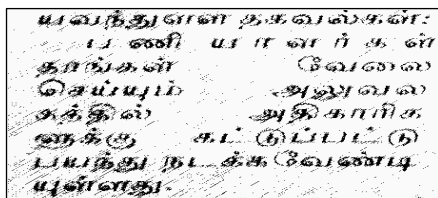
The following observations have been made on Backside Text visible documents in Tamil script.

- (a) This problem is quite common when papers are thin or poor in quality.
- (b) Deep dark printing on pages resulting in bleed through.
- (c) Due to the partial appearance of backside characters the original characters on the front are misunderstood.
- (d) Binarization processes of these documents have registered a lot of noise pixels.
- (e) All segmentation processes prove to be failure in front of this distortion at most of the places.

5. Faxed Documents

Fax machines are considered one of the major sources of text distortion, when they create the problems in recognition of their own. This distortion is very apparently visible in the form of spurious point noise and ragged edges. Fax involving light printing, produces a huge number of broken characters, few touching characters and sometimes only a few pixels of characters. Distortions of fax are listed out as salt and pepper noise, the thickening or the partial omission of figures cause inappropriate threshold sensor outputs, which adds random noise and ill balanced bias that should be overcome.

Figure 5: Faxed Document in Tamil Script



However, generally it is difficult to restore distorted fax images of fax, as almost all the gray scale information of the output image is lost when using the binarization process. Even the human beings sometimes do not recognize distorted images.

The observation read as the list follows:

- (a) The width of the stroke is not constant over the document.
- (b) Entire document is found with all types of distortions at all the zones.
- (c) The quality of the fax document depends upon the fax machine.

6. Type Written Documents

Typewritten documents also encounter the problem of distortion. Typewriters are widely used in Government offices of India. As far as Tamil scripts are concerned, 80% of official works are processed by means of type written documents only.

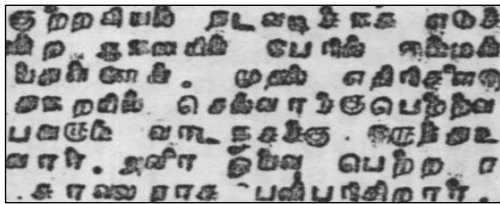


Figure 6: Typewritten Document in Tamil Script

A sequence of statistical analysis of typewritten characters has brought out the following observation.

- (a) The middle zone is terribly influenced by the distortion in the form of merged characters.
- (b) Lower zone characters, almost every time merge with the previous upper zone characters complicating the segmentation of lower zone from the middle zone.
- (c) Unequal spacing between lines, words and characters has been observed.
- (d) There is a significant change in the shape of upper zone and lower zone characters.
- (e) Some times the characters are broken into many parts placed at various base lines. Usually most of the difficult algorithms have been designed on the basis of headline and baseline.
- (f) Extra force applied during typing leads to character distortion of heavy printing.
- (g) The technical inability of the typewriter with fixed width or variable grid produces characters of the same horizontal shape irrespective of the actual shape of the characters.

Conclusion

On the whole, this study has shouldered the task of venturing into the treatment of distorted characters of Tamil scripts. Aiming at the maximization of the recovery rate of damaged or dilapidated documents of various Indian languages with special highlights on Tamil scripts was covered under different headings viz., touching characters, broken characters, heavily printed documents, faxed documents, typewritten documents and the like. This study not only probes into the character recognition process that deals with so many problems of distorted Tamil characters, the kinds of distortions on it and the possible solutions

to overcome those stumbling blocks, but also wide spreads ample of opportunities and scope for further researches in the same field of character recognition. Though researches enlist a range of character distorters like spray marks, curved base lines, blurred images, presence of punctuation marks and so on, this study pacts the treatment of only a handful initiation to the researches in the field of character recognition of Indian scripts in general and Tamil scripts in particular.

Reference:

1. Y. Li, D. Lopresti, G. Nagy and A. Tomkins, "Validation of image defect models for optical character recognition", *IEEE Transactions on PAMI*, Vol. 18(2), pp. 99-108, 1996.
2. H. S. Baird, "The state of the art of document image degradation modeling", invited talk, in the Proceedings of Int., Workshop on Document Analysis Systems, Rio de Janeiro, Brazil, pp. 10-13, 2000.
3. S. W. Lee, D. J. Lee and H. S. Park, "A new methodology for gray-scale character segmentation and recognition", *IEEE Trans., on PAMI*, Vol. 18(10), pp. 1045- 1050, 1996.
4. C. B. Bose and S. S. Kuo, "Connected and degraded text recognition using hidden markov model", *Pattern Recognition*, Vol. 27(10), pp. 1345-1363, 1994.
5. S. Tsujimoto and H. Asada, "Resolving ambiguity in segmenting touching characters," 1st Int., Conf., on Document Analysis and Recognition, pp. 701-709, Saint-Marlo, France, Sept. 1991.
6. R.G. Casey and G. Nagy, "Recursive segmentation and classification of composite character patterns," *Proc. 6th Int., Conf., on Pattern Recognition*, pp. 1023-1026, Munich, 1982.
7. T. Hong, Degraded Text Recognition using Visual and Linguistic Context, Ph. D. thesis, Computer Science Dept., of SUNY at Buffalo, 1995.
8. S. Kahan, T. Pavlidis and H. S. Baird, "On the recognition of printed characters of any font and size", *IEEE Trans. Pattern Anal. Mach. Intell.* 9, 274 288 (March 1987).
9. Y. Lu, "On the segmentation of touching characters," in *Proc. Int. Conf. Document Anal. Recognition*, Tsukuba Science City, Japan, 1993, pp. 440-443.
10. U. Garain and B. B. Chaudhuri, "Compound character recognition by run number based metric distance", *SPIE Proc.*, Vol. 3305, pp. 90-97, 1998.
11. B. B. Chaudhuri, U. Pal and M. Mitra, "Automatic recognition of printed Oriya script", in the Proceedings of 6th ICDAR, pp. 795-799, 2001.
12. M. K. Jindal, G. S. Lehal and R. K. Sharma, "Segmentation of touching characters of Indian scripts-an overview", in the proceedings of National Conference on Recent Advances and Future Trends in IT (RAFIT 2005), Punjabi University Patiala, pp. 74-77, 2005.
13. G. S. Lehal and C. Singh, "Text segmentation of machine-printed Gurmukhi script", *Document Recognition and Retrieval VIII*, proceedings SPIE, USA, Vol. 4307, pp. 223-231, 2001.
14. G. S. Lehal and C. Singh, "A technique for segmentation of Gurmukhi text", *Computer Analysis of Images and Patterns*, Proc. CAIP 2001, W. Skarbek (Ed.), Lecture Notes in Computer Science, Vol. 2124, Springer-Verlag, Germany, pp. 191-200, 2001.
15. V. Bansal, Integrating Knowledge Sources in Devanagari Text Recognition, Ph. D. thesis, IIT Kanpur, India, 1999.
16. R.M.K.Sinha and H.Mahabala, "Machine recognition of Devanagari script", *IEEE Trans. Syst. Man Cybern.* Vol. 9, 1979.
17. A. Whichello and H. Yan, Linking broken character borders with variable sized masks to improve recognition, *PR*, vol. 29, pp. 1429. 1435, August 1996.

18. M. Bern and D. Goldberg, Scanner-model-based document image improvement, in ICIP00, pp. Vol II: 582-585, 2000.
19. T. Akiyama, N. Miyamoto, M. Oguro, and K. Ogura, Faxed document image restoration method based on local pixel patterns, in SPIE98, vol. 3305, pp. 253-262, Apr. 1998.
20. Y. Lu, B. Haist, L. Harmon, J. Trenkle and R. Vogli, "An accurate and efficient system for segmenting machine-printed text", U.S.Postal Service 5th Advan., Tech., Conf., Washington, Vol.3, pp.A93-A105, 1992.
21. O. Nakamura, M.Ujiie, N.Okamoto and T. Minami, "A character segmentation algorithm for mixed-mode communication", Trans. IEICE, (D) 167-D, 11, pp. 1277- 1285, 1984.
22. N. Okamoto, O. Nakamura and T. Minami, "Character segmentation for mixed-mode communication", IFIP'83, pp.681-685, 1983.
23. B.A.Yanikoglu, "Pitch - based segmentation and recognition of dot-matrix text", Int., Journal of Doc.,t Analysis and Recognition (IJ DAR), Vol.3, pp.34- 39, 2000.
24. M. Droettboom, "Correcting broken characters in the recognition of historical printed documents", in the Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital libraries (JCDL), Houston, Texas, USA, pp. 364-366, 2003.
25. Stephen V. Rice, George Nagy and Thomas A. Nartker, Optical Character Recognition: An Illust., Guide to the Frontier, Kluwer Academic Pub., 1999.
26. G. Leedham, S. Varma, A. Patankar, and V. Govindaraju, "Separating text and background in degraded document images-a comparison of global thresholding techniques for multi-stage thresholding," Proc. 8th IWFHR, Aug-2002, pp.244-249.
27. Anna Tonazzini, Emanuele Salerno, and Luigi Bedini, "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique," IJDAR, vol. 10, no. 1, pp.17-25, 2007.
28. E. Dubois and A. Pathak, "Reduction of bleed-through in scanned manuscript documents," in Proc. IS&T Image Processing, Image Quality, Image Capture Systems Conference (PICS2001), Montreal, Canada, April 2001, pp. 177-180.
29. Google, Book Search Dataset, Version v edition, 2007.
30. Gang Zi and D. Doermann, "Document image ground truth generation from electronic text," in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th Int.,Conference on, D. Doermann, Ed., 2004, vol. 2, pp. 663-666 Vol.2.
31. Gang Zi, "Ground truth generation and document image degradation," Tech. Rep. LAMP-TR-121/CARTR- 1008/CS-TR-4699/UMIACS-TR-2005-08, University of Maryland, College Park, 2005.
32. Dan S. Bloomberg, "Determining the Resolution of Scanned Document Images", Presented at IS&T/SPIE EI'99, Conf.,ce 3651, Doc., Recognition and Retrieval VI, Jan 26-28, San Jose, CA.
33. M. Oguro, T. Akiyama and K. Ogura, "Faxed document image restoration using gray level representation", in the Proc., of 4th ICDAR, Vol. 2, pp. 679-683, 1997.
34. T.R. Randolph and M.J.T. Smith, "Enhancement of fax documents using a binary angular representation", in the Proc., of Int., l Symp., on Intelligent Multimedia, Video and Speech Processing, Hong Kong, pp.125-128, 2001.
35. J. D. Hobby and T. K. Ho, "Enhancing degraded documents images via bitmap clustering and averaging", in the Proc., 4th ICDAR, pp.394-400, 1997.
36. P. Natarajan, I. Bazzi, Z. Lu, J. Makhoul and R. Scwhartz, "Robust OCR of degraded documents", in the Proceedings of 5th ICDAR, pp. 357-361, 1999.

37. M. Cannon, J. Hochberg and P. Kelly, "QUARC: a remarkably effective method for Increasing the OCR accuracy of degraded typewritten documents", in the Proc., of the 1999 Symposium on Doc., Image Understanding Tech., (SDIUT'99), Annapolis, MD, pp. 154-158, May 1999.
38. M. Cannon, J. Hochberg and P. Kelly, "Quality assessment and restoration of typewritten document images", IJDAR, Vol. 2(2-3), pp. 80-89, 1999.
39. C. Rodriguez, J. Muguerza, M. Navarro, A. Zarate, J.I. Martín and J.M. Perez, "Segmentation of low-quality typewritten digits", in the Proc., 14th ICPR, pp. 1106.

A High Accuracy Phone Recognition System for Tamil

C. P. Santhosh Kumar

Amrita Vishwa Vidyapeetham, Coimbatore
email: cs_kumar@cb.amrita.edu

N. Deiva Sundaram

Madras University, Chennai

Abstract

Phone recognition systems are used in many applications such as automatic segmentation of speech, keyword spotting, automatic language identification using phonotactic approach, and speaker identification. Phone recognition accuracy and the accuracy of the application developed are highly correlated. In this paper, we present the details of a high accuracy phone recognition system developed for Tamil. We use a hybrid hidden Markov model – neural network to implement the decoder. It was seen that for moderate data sizes, the system outperforms the hidden Markov model – Gaussian mixture model implementations.

Introduction

High accuracy phone recognition systems are very useful for many applications. They can be helpful for phonetic labeling of speech; speech data collected with the word level transcriptions without any segmentation information can be converted to phonetic transcriptions with segmentation information. If this segmentation information has to be generated manually, we will need an experienced phonetician, and doing this manually is not cost effective. Further, due to the co-articulation effects, the segmentation boundaries thus obtained will not be precise. Under such conditions, a consistent phone boundary is the best choice, and a machine assisted segmentation and labeling using a phone recognizer can be more effective and cost effective. Phone recognition systems are also used in automatic keyword spotting systems using phone lattices, language identification using phonotactic approach, and speaker identification. Performance of these systems is highly correlated with the phone recognition accuracy and therefore any effort to enhance the phone recognition accuracy is likely to enhance the performance of the application.

Hidden Markov model –Gaussian mixture models (HMM-GMM) are known for their performance in the development of speech recognition systems, and in such systems Gaussian mixture models are used to model the probability distributions. In speech recognition systems, word n-gram language models and phone n-gram language models can be used to enhance the recognition accuracy. However, for moderate size databases hidden Markov model – neural network structures[1,2,3,4] have been found to offer better performance compared to the HMM-GMM systems, and for large databases both offer similar performances, but HMM-NN systems has less complexity.

Segments of speech (frames) represented in terms of their probabilistic similarity to the phones of any language are often referred as *probabilistic features (PF)*[1, 2]. Temporal patterns (TRAPS)[1] of log energy from critical bands is a way to derive PFs. In [1, 2], several band-conditioned classifiers are used to derive

PF to be merged by a neural network to estimate the phone posteriors. The use of band conditioned phone posteriors as temporal features was studied in [3]. In this work, they used the hidden layer activation outputs, and is known as hidden activation TRAPS (HATS). In [4], a simplified system offering better results, and requiring less training data due to the splitting of TRAPS features to left and right context has become popular since [4] due to its simplicity and ability to reduce the amount of training data required for similar performances. In this work, we use the implementation in [4] to implement a high accuracy phone recognizer for Tamil.

Keyword spotting systems (KWS)[5,6] are useful for archiving and indexing audio/video documents to be searched using keywords, searching for telephonic conversations that contain words/phrases that are potentially dangerous to the national security, and thereby derive intelligence information. Large vocabulary continuous speech recognition (LVCSR) system based KWS are very popular for their better performance. However, they have many limitations to be applied directly to the Indian environment. To develop an LVCSR system for a language, we need a large labeled speech database to train the acoustic models, and to the best of our knowledge, such a database is not yet available for Tamil. In the Indian context, we tend to use words/phrases across languages, and is fairly common to mix words/phrases from English in native language conversations. LVCSR systems cannot handle multiple languages at the same time, as it needs a language model to enhance the word recognition accuracy, and training a language model catering for multiple languages is practically impossible for many reasons. In this work, we present the details of a KWS systems developed using the lattices [5] generated by the HMM-NN phone recognizer. Another approach that is widely popular uses lattices generated by a phone recognizer [5, 6]. This approach has the advantage that it can be easily ported to multilingual environments, unlike the LVCSR based approach. [5] gives a comparison of different KWS approaches.

HMM-NN Phone Recognizer

In the hybrid HMM-NN system, critical band energies are obtained in the conventional way [1, 2, 3, 4], Speech signal is divided into 25 ms long frames with 10 ms shift. The Mel filter-bank is emulated by triangular weighting of FFT-derived short-term spectrum to obtain short-term critical-band logarithmic spectral densities. TRAPS feature vector describes a segment of temporal evolution of critical band spectral densities within a single critical band. The central point is the current frame and 15 frames from the past make a left context (LC) feature and similarly 15 frames from the future make the right context (RC) feature vector. Further, the LC and RC features are windowed to make the transition between the successive frames smooth. Triangular window was used for this purpose. Subsequently, these feature vectors are processed for dimensionality reduction. We used discrete Cosine Transform (DCT) for its simplicity to reduce the 16 dimensional LC and RC feature sizes to 11 [1], [2], [3], [4]. To further enhance the accuracy of the system, we concatenated 15 critical band features for LC and RC to generate input to two separate LC and RC neural networks. Outputs of these classifiers are subsequently merged together using another neural net. Outputs of all neural networks represent phone state posterior probabilities, and phone models have three states each. Details of the implementation can be found in [3], [4] Fig. 1 illustrates the schematic diagram of the implementation of the LC-RC HMM-NN phone recognizer.

Fig. 1 – Hybrid HMM-NN LC-RC phone recognizer [4]

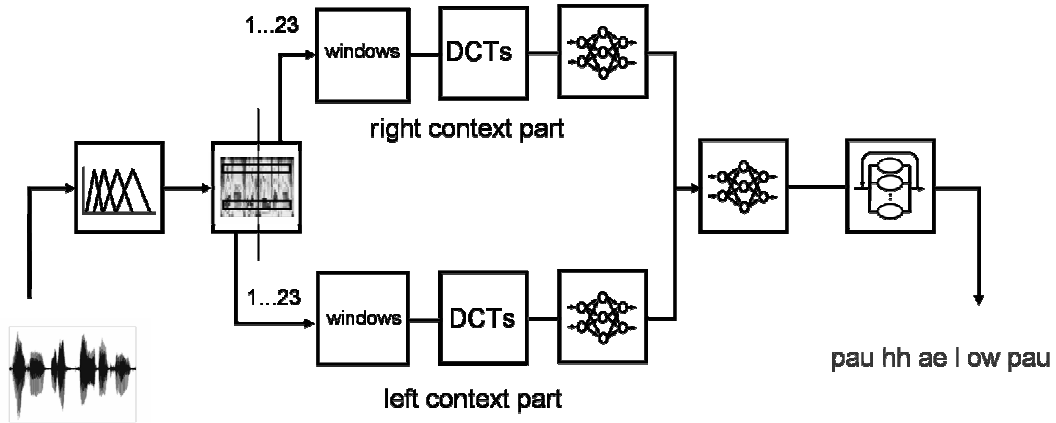


Table 1 – Tamil Vowels used in the recognizer with illustrating examples

Phone (ASCII)	Alphabet – Example Transcription	Phone (ASCII)	Alphabet – Example Transcription
ax	அ - அம்மா /ax m h m aa/	ay	ஏ - எட்டு /ae t h t U/
aa	ஆ - ஆடு /aa d U/	ey	ஏ - ஏணி /ey nn ih/
ih	இ - இலை / ih l ax y/	ao	ஒ - ஒன்று /ao n dr U/
iy	ஈ - ஈகை /iy hh ax y/	ow	ஓ - ஓம் /ow m/
uh	உ - உணவு /uh nn ax v U/	aw	ஔ - ஔவை /aw v ax y/
U	உ - தப்பு /tx ax p h p U/		
uw	உள - உளர் /uw r/		

Table 2 – Tamil consonants used in the recognizer with illustrating examples

Phone (ASCII)	Alphabet – Example Transcription	Phone (ASCII)	Alphabet – Example Transcription
k	க் - கல்வி /k ax l v ih/	dh	த் - பந்தம் /p ax nd dh ax n/
g	க் - பங்கு /p ax ng g U/	nd	ந் - பந்து /p ax nd dh U/
ng	ங் - தங்கம் /t ax ng g ax m/	p	ப் - பாடம் /p aa d ax m/
ch	ச் - காட்சி /k aa t ch ih/	b	ப் - அம்பு /ax m b U/
jh	ஜ் - ஜாதி /j aa tx ih/	p	ப் - லாபம் /l aa p ax m/
ny	ஞ் - ஞானம் /ny aa n ax m/	m	ம் - கரும்பு /k ax r uh m b U/
t	ட் - அட்டை /ax t ey/	y	ய் - யமுனை /y ax m uh n ax y/
d	ட் - கடன் /k ax d ax n/	r	ர் - மரம் /m ax r ax m/
nn	மணி /m ax nn ih/	tr	ற் - ஒற்றன் /ao t tr ax n/
ddn	ண் - கண் /k ax ddn/	dr	ற் - கன்று /k ax n dr U/
th	த் - தட்டு /th ax t h t U/	rr	ற் - அறம் /ax rr ax m/
Th	த் - மதி /m ax Th ih/	l	ல் - மாலை /m aa l ax y/
h	பட்டம் /p ah t h t ah m/	ll	ள் - உள்ளம் /uh ll h ll ax m/
sh	ஷ் - விஷம் /v ih sh ax m/	lzh	ழ் - பழம் /p ax lzh ax m/
s	ச் - சங்கு /s ax ng g U/	v	வ் - வயது /v ax y ax dh U/
hh	க் - பகல் /p ax hh ax l/	n	ன் - நான் /nd aa n/

Lattice based KWS system

Lattices [7] are a way of storing output of a recognizer as an oriented acyclic graph in which each node represent a symbol (phone) and each link represent time boundaries of the phone/word at the end of the link. Fig. 2 shows an example lattice generated by a phone recognizer. It may be seen that at any instant in time, there are multiple possibilities available when the keywords are to be searched. Searching in lattices provides better results than searching in phone strings [5] as it holds several hypotheses in parallel. We follow a lattice based approach in this work [5,6]. Phone lattices were generated from phone posteriori probabilities. In our experiments phone insertion penalty was set to zero for lattice generation to minimize deletion of phones. Fig. 3 shows the schematic representation of the KWS system implemented in this paper, and the likelihood of the keywords are compared with the likelihoods of the hypothesis generated by the background filler phone models working in a parallel loop to decide the presence or absence of a keyword in a continuously spoken sentence.

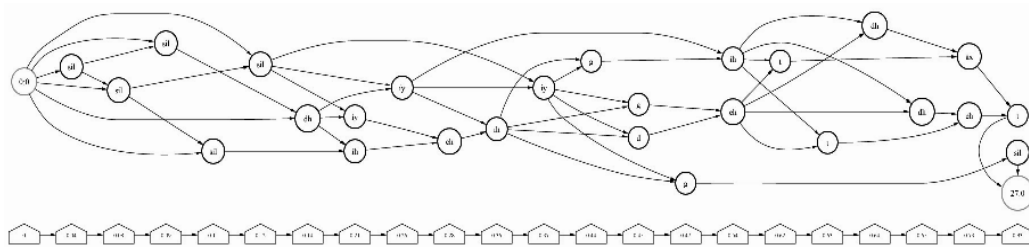


Fig. 2 – An example of a lattice generated by the phone recognizer

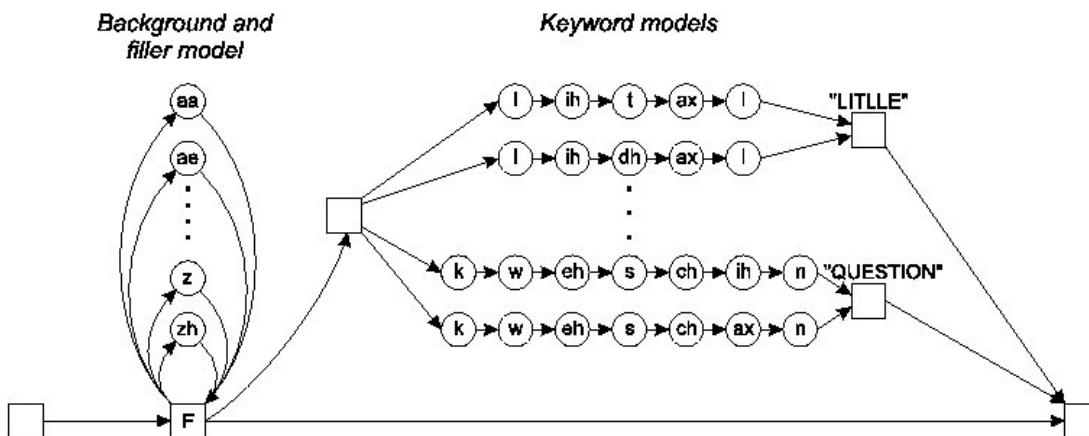


Fig. 3 – A Schematic description of a KWS system

Experiments and Results

We used 1.2 hours of telephone quality speech sampled at 8 kHz for our experiments. In the HMM-NN systems, we used 300 neurons for modeling the probability distribution. For the HMM-GMM systems, we used 16 Gaussians to model the probability distributions, and the models are trained using maximum mutual information criteria using the HTK toolkit [7]. The size of the neural network and the GMMs was chosen to match the limited training data available.

HMM-GMM	HMM-NN
51.87	56.62

Table 3 – Comparison of the performance of the HMM-GMM and HMM-NN phone recognizers in per cent for Tamil

KWS systems are evaluated using Figure-of-Merit (FOM)[6], which is the average of correct detections per 1, 2, . . . 10 false alarms per hour. We used the most frequently occurring 60 words to benchmark our system. The KWS system developed in this using the HMM-NN phone recognizer gave an FOM of 66.25 per cent for the most frequently used 60 words in the test set.

References

1. H. Hermansky, and S. Sharma, TRAPS - classifiers of temporal patterns, Proc. ICSLP, Sydney, Nov. 1998
2. H. Hermansky, D.P.W. Ellis, and S. Sharma, Tandem Connectionist feature extraction for conventional HMM systems, Proc. ICASSP 2000, Turkey, 2000.
3. B. Chen, Q. Zhu, and N.Morgan, Learning long term temporal features in LVCSR using neural networks, Proc. ICSLP, Jeju Island, Oct. 2004.
4. P. Schwarz, P. Matejka, and J. Cernocky: Towards lower error rates in phoneme recognition, in *Proc. TSD 2004*, Brno, Czech Republic, 2004
5. I. Szoke and P Schwarz and P. Matejka and L. Burget and M. Karafi'at and M. Fapso and J. Cernock'y, Comparison of Keyword Spotting Approaches for Informal Continuous Speech, Interspeech 2005 - Eurospeech, Lisboa, Portugal, Sep. 2005, pp. 633-636.
6. M. Saraclar and R. Sproat, Lattice-Based Search for Spoken Utterance Retrieval, Human Language Technology Conference of the North American Chapter of the Association for Computational linguistics (hlt-naacl2004), Boston, Massachusetts, USA, May, 2004.
7. CUED, HTK toolkit, <http://htk.eng.cam.ac.uk/>

Face Waves: 2D - Facial Expressions Based on Tamil Emotion Descriptors

Sabitha.Tammaneni, Madhan Karky

tammaneni@yahoo.com, madhankarky@gmail.com

Department of Computer Science & Engineering

College of Engineering Guindy

Anna University

Abstract

This paper aims at recognizing human emotions from textual descriptors and expressing the emotion in a 2 dimensional computer generated face for the Face Waves framework. The main objective of this paper is to map facial expressions for the basic six emotions (Love, Joy, Fear, Anger, Surprise, and Sadness), 27 second level derived emotions, and a third level 79 emotions. Apart from hierarchically classifying these 112 emotions in Tamil, this paper presents an object oriented emotion model, where the emotions at lower levels inherit properties from emotions at higher levels. The paper also describes an 'expression index' built for efficiently storing and retrieving the facial features for corresponding expressions for every emotion. The 'expression index' plays a vital role in mapping the expression on a computer-generated face with time efficiency. Discussing the results of mapping and efficiency of the index, the paper concludes with open questions and future enhancements.

Introduction

Since Charles Darwin's early work on 'The Expression of Emotions in Man and Animals', there have been various proposals to classify human emotions. A few researches have been carried out in mapping expressions to emotions. The expressions to emotion mapping cannot be generalized as every individual has his or her own way of expressing various emotions. In many cases it will not be possible for one to distinguish extreme happiness from extreme sadness just by observing facial expression of a person. In this paper we present a system that can identify emotions from textual descriptions and try to map a corresponding expression on a 2D computer generated face.

The primary aim of this paper is to present an object model for all possible human emotions and their corresponding Tamil textual descriptors. The secondary aim is to map the action units of various parts of a human face to express the emotion. To best of our knowledge, this is the first time that such a work is presented in Tamil and first time an object-oriented model is proposed for representing emotions and their expressions.

This paper is organized into five sections. The second section provides some background information and discusses literature related to this work. The third section presents the classification tree for human emotions in Tamil. The fourth section describes the system design and explains the action units for every expression. The fifth section presents the results and discusses the advantages of an expression index based on the object-oriented emotion tree to increase the time efficiency of expressing emotions on a 2D face for a text-to-video system.

Background

“Facial expression” has drawn interest to a few researchers around the world. Most of the research focus on expressing emotions in a 3D face. Paul Ekman and Wallace F introduced a system called FACS, where each facial expression is described in terms of Action Units (AU’s) [1]. Irfan A. Essa and Alex P. Pentland implemented a computer vision system which is developed by using mathematical formulation which is used for detailed analysis of facial expressions[3]. Prem Kalra ,Angelo Mangili,Nadia Magnenat Thalmann and Daniel Thalmann made every expression from one or more MPA's(Minimum Perceptible Action),One or several simulated muscle actions constitute a MPA [2]. Praseeda Lekshmi.V And Dr.M.Sasikumar used multicast Support Vector Machine for Classification of different kinds of facial expressions belonging to the face image they also used Gabor filters for image processing[4]. Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi developed a deformable muscle-based face model that tracks some FCP's in the real face image sequences and shows the same expressions[5]. Robert Plutchik created a wheel of emotions, which forms the base to this paper[6]. The emotions provided in English are translated in Tamil to present the emotion tree.

Classification of Human Emotions

As explained in section 1, there have been various proposals to classify human emotions. Many classifications do not recognize love as a basic human emotion. We have achieved our classification of emotions based on Robert Plutchik’s classification[6]. The first level of tree is provided in figure 1. Happiness, Fear, Love, Sadness, Anger and Surprise form the first level of emotions. All other emotions are derived from these basic emotions according to Robert.

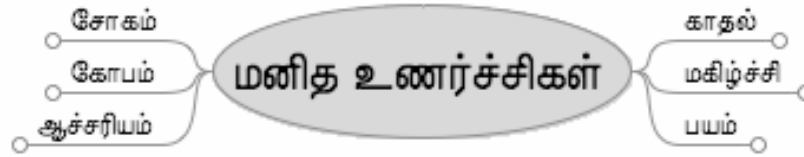


Figure 1 : Emotion Tree : Level 1

Each of these basic emotions in the first level of tree has sub-emotions as shown in figure 2. Figure 2 gives the second level emotions for two of the basic emotions happiness and sadness.



Figure 2 : Emotion Tree : Level 2

The classification of these emotions runs down to a third level where each of the sub-emotions have sub-sub-emotions with the same base emotion. Figure 3 gives the third level emotions for happiness and sadness. The sub emotion shame(avamaanam) gives raise to four third level emotions. Similarly the sub-emotion cheerfulness(uRsaagam) gives raise to eight third level emotions.

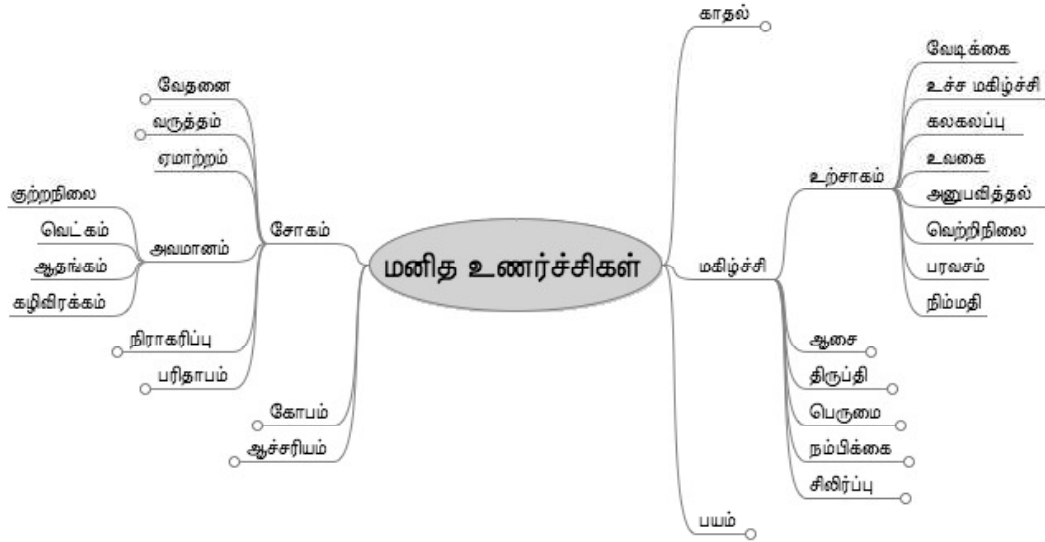


Figure 3 : Emotion Tree : Level 3

The complete tree consisting of 112 emotions is not provided in this paper owing to space considerations.

System Design

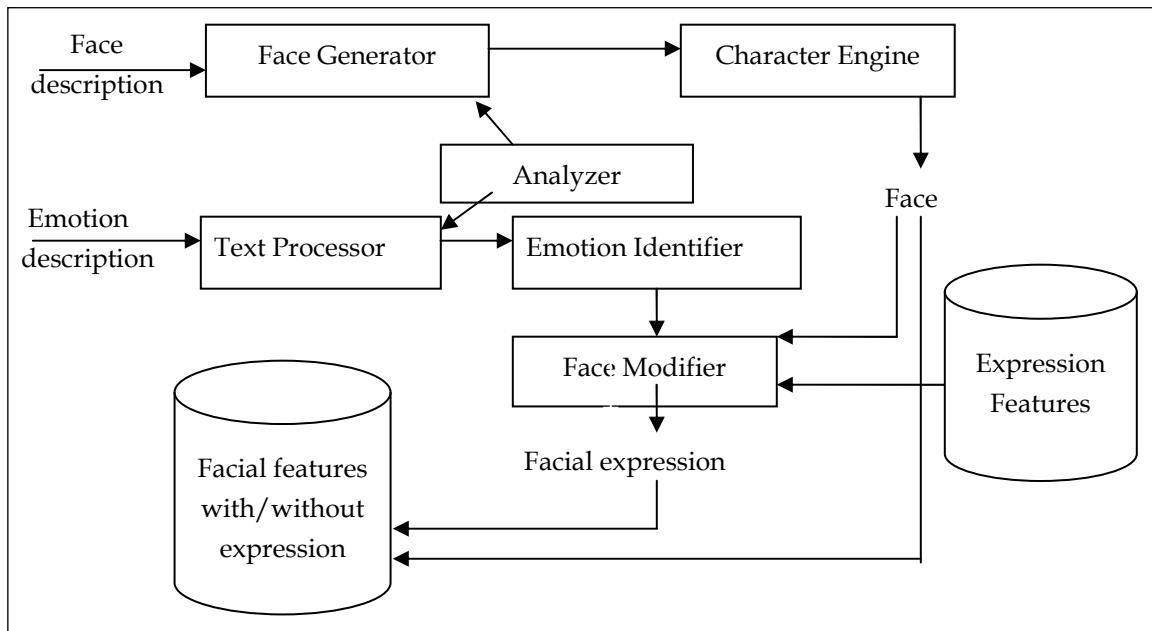


Figure 4 : Emotion Processor System Design

The design of our emotion processor is given in figure 4. Face Generator will accept the input (face description) from the user, which will analyze the input and then distributes the corresponding characteristics for each part of the face. The character Engine will construct the Human face, which is a 2D computer generated face.

The Text Processor processes a given text document for emotional descriptors. The text lines corresponding to the emotional descriptors are analysed using a morphological analyzer and the output of the analyzer is fed to the Emotion Identifier. Emotion identifier using the emotion model explained in section 3, identifies the emotion and sends the emotion tag to Face Modifier. The Face Modifier module uses an expression index to modify every part of the face according to the expression. Each part of face is modeled as an object and the face modifier based on the emotion tag controls the attributes of the object. Thus the emotion is mapped to the face that comes as input from the character engine as a set of feature changes termed as an expression of the emotion.

The modified face can now be stored as text. in a database for efficient retrieval for the Text-To-Video system. It is to be noted that the Face Generator may generate any type of a face based on the descriptions it gets. The pre-generated face is given as input to the Face modifier along with the emotion descriptors.

Emotion to Expression

We use Ekman's FACS Action Units(AU)[11] to describe the transformations on a 2D computer generated face. The action units specify pull, raise like actions for every part of the face. Those Aus corresponding to 2D face alone have been used to represent the emotion on the face as an expression. The following are examples of two basic emotions and their corresponding action units.

"மகிழ்ச்சி"(Joy):

AU	Description
12	Lip corner puller
5	Upper lid raiser

Table 1 : Action Units for Joy

"பயம்"(Fear):

AU	Description
1	Inner brow raiser
2	Outer brow raiser
5	Upper lid raiser
7	Lid tightener
20	Lip stretcher

Table 2 : Action Units for fear

Emotions in the second and third levels of the emotion tree inherit the action units from their parent emotion and have their own action units along with the inherited units.

Results

A simple face generated by the Face Generator module and a few samples of the faces modified by our Face Modifier for different emotions are given in figure 5. The features of the face are given as input to the Face Modifier. Emotion descriptors can be natural language text describing the state of the person.

அவன் சோகமாக காணப்பட்டான்

He looked sad

அவன் அப்போது பயந்துபோய் கிடந்தான்

He looked scared at that moment

The above given sentences can be examples of emotion descriptors. Our analyzer breaks the words to identify the root and when the emotion is identified, the corresponding features are retrieved from the expression index. The features are applied over the 2D face that was received as input for modification.

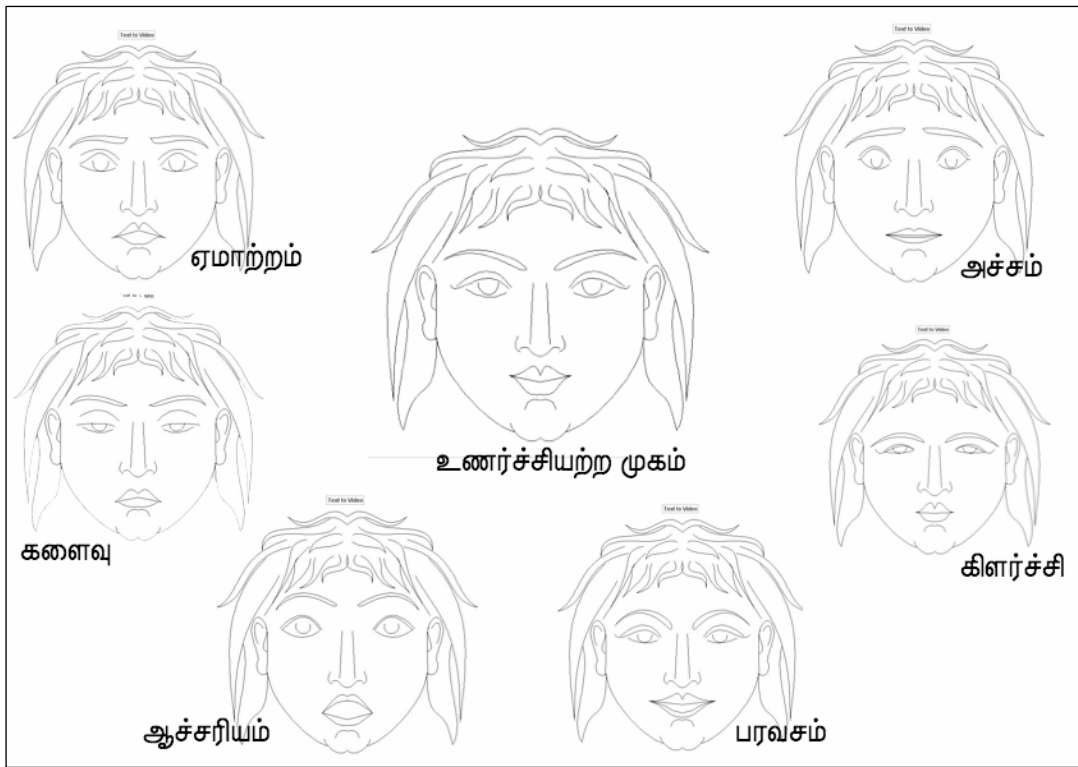


Figure 5 : Original Face and Modified Faces

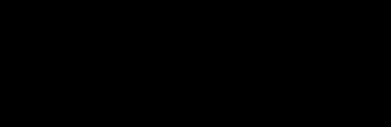
The face in the center of figure 5 is the original face generated by our face generator surrounded by the corresponding faces modified for six different emotions. Similarly any generated face is modified for one of the 112 emotional descriptors.

Conclusion and Future Work

This paper proposes an object-oriented model for representing emotions hierarchically and proposes a system for applying emotions on a 2D computer generated face using Ekman's FACS action units. Improving the index for minimizing the number of changes from original face to modified face will be our immediate research that follows. Such an index will improve the time efficiency of creating text-to-video. Applying these emotions for a 3D face can aid to generating 3D animation videos.

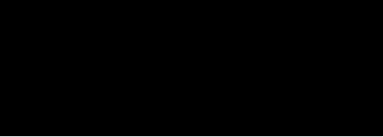
References

1. Ekman.P,W. V. Friesen, "*The facial action coding system: A technique for measurement of facial movement*". 1978.
2. Prem Kalra ,Angelo Mangili,Nadia Magnenat Thalmann and Daniel Thalmann, "*3D interactive free form deformations for facial expressions*", First International Conference on Computational Graphics and Visualization Techniques, Sesimbra, Portugal, 1993.
3. Irfan A. Essa, Alex P. Pentland, "*Coding, Analysis, Interpretation, and Recognition of Facial Expressions*," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.
4. Praseeda Lekshmi.V and Dr.M.Sasikumar, "*Analysis of Facial Expression using Gabor and SVM*", International Journal of Recent Trends in Engineering 2009.
5. Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi, "*Facial Expressions Animation and Lip Tracking Using Facial Characteristic Points and Deformable Model*", International Journal of Information Technology , 2005
6. List of Human Emotions, http://en.wikipedia.org/wiki/List_of_emotions, Last Accessed 23/04/2010



6

**தமிழ் மின் தரவு மற்றும்
மின்னகராதிகள்**



மலாயாப் பல்கலைக்கழக மின்னியல் நூலகம்

Ilangkumaran S, O Sivanadhan
s.ilangkumaran@gmail.com

மின்னியல் நூலகம் என்பது, தற்போது வழக்கில் உள்ள அச்ச வடிவம் அல்லது மைக்ரோ•பிலிம் வடிவத்தில் உள்ள புத்தகங்களை அல்லது அவற்றுள் ஒரு பகுதியை மாற்று வழியாகவோ, மறு பிரதியாகவோ, துணைப்பொருளாகவோ மின்னியல் வடிவில் பாதுகாத்து வைக்கும் நூலகமாகும். தரவுகளைப் பயன்ட்டாளர்கள் எனிய முறையில் பயன்படுத்திக் கொள்ளும் வண்ணம் இந்நூலகம் முறையாக வரிசைப்படுத்தப்பட்ட பல்லாடகச் சாதனங்களைக் கொண்டிருக்கும்.

நூலகத் தகவல் சாதகங்களைக் கணினி மயமாக்குதல் இன்றைய காலகட்டத்தில் அவசியமாவதோடு, பொதுமக்களுக்குத் தகவல்களைக் கொண்டு சேர்ப்பதற்குச் சிறந்த வழியாக உள்ளது. அதோடு, நூலகச் சேவையை விளம்பரப்படுத்தும் ஒரு யுக்தியாகவும் இத்திட்டம் செயல்படுகின்றது. ஒரு நூலகம் தனித்தன்மையோடு விளங்க வேண்டும் என்ற நோக்கமும், பொது மக்களின் பல்வேறு தகவல் தேவைகளைப் பூர்த்தி செய்யும் எண்ணமே மலாயாப் பல்கலைக்கழக நூலகம் எனிய முறையில் பலதரப்பட்ட தகவல்களை அறிமுகப்படுத்த தூண்டுகோலாக அமைந்தது.

மலாயாப் பல்கலைக்கழக நூலகம் தொடங்கிய காலத்தில் பதிவு புத்தகம் மற்றும் பதிவு அட்டை வழியே தகவல் சாதனங்கள் பாதுகாக்கப்படும் தேடப்படும் வந்தன. இருப்பினும் இன்றைய தகவல் தொழில்நுட்ப வளர்ச்சியில் இருந்து, மலாயாப் பல்கலைக்கழக நூலகமும் பின் தங்கி இருக்காது, தகவல் பரிமாற்றத்திற்கு மின்னியல் நூலகத்தின் வழி அதன் தரத்தை உயர்த்தி வருகின்றது. தானியங்கி முறையின் வழி நூலக அலுவல்களை விரைவாகவும், முழுப் பயன் அடையும் விதமாகவும், சிறப்பாகவும் விரிவுபடுத்தி செயல்படுத்தவுள்ளது.

மலாயாப் பல்கலைக்கழகத்தில், மலாயாப் பல்கலைக்கழக நூலகமும், கணினி அறிவியல் மற்றும் தகவல் தொழில்நுட்பத் துறையும் இணைந்து மின்னியல் நூலக மேம்பாட்டினைச் செய்து வருகின்றன.

மின்னியல் நூலகம் சிறந்த வழிமுறைகள்: காங்கிரஸ் உலக மின்னியல் நூலகம்.

உலக மின்னியல் நூலகம் என்பது, UNESCO மற்றும் அமெரிக்க காங்கிரஸ் நூலகம் இணைந்து வழிநடத்தும் அனைத்துலக மின்னியல் நூலகமாகும். இந்த நூலகத்தின் வாயிலாக பின்வரும் சேவைகள் அனைத்துலக மக்களுக்கு வழங்கப்படுகின்றன :

- அனைத்துலக மற்றும் பல்வேறு கலாச்சாரப் புரிந்துணர்வை வெளிப்படுத்துதல்
- கலாச்சாரம் சார்ந்த தகவல்களை இணையத்தில் அதிகரித்தல்.
- கல்வியாளர்கள், மாணவர்கள் மற்றும் பொது மக்களுக்குத் தேவையான தகவல்களை வழங்குவதோடு இணை இயக்கங்களின் தரத்தை மேம்படுத்தி, நாடுகளுக்கு இடையிலான தொழில் நுட்பப் பிரிவினைகளைக் குறைத்தல்.
- இலவசப் பன்மொழி வடிவிலான உலகளாவிய கலாச்சாரத்தின் முக்கியத் தகவல்களான எழுத்து வடிவம், வரைபடம், அரிதான நூல்கள், இசை, பதிவுகள், திரைப்படங்கள், அச்சுகள், புகைப்படங்கள், கட்டிடக்கலை வரைபடங்கள், முக்கியத்துவம் வாய்ந்த கலாச்சார பதிவுகள் போன்றவற்றை இணையத்தில் கிடைக்க வழி செய்தல்.

இந்த நூலகத்துக்கு உலகளாவிய ரீதியில் 30-க்கும் மேற்பட்ட தேசிய நூலகங்களோடும் கல்வி நூலகங்களோடும் ஒப்பந்த இணைப்பு உள்ளது. மலாயாப் பல்கலைக்கழக நூலகமும் இந்த காங்கிரஸ் உலக மின்னியல் நூலகச் செயல்பாடுகளையே பின்பற்றி வருகின்றது.

மின்னியல் நூலகப் பயன்பாட்டின் நோக்கம்:

அ) நூலகத்திலிருந்து பெற விழையும் தகவல்களை, நூலகத்துக்குச் செல்லாமலேயே

எங்கிருந்து வேண்டுமானாலும் எளிதாக கையாளலாம்

ஆ) நூலகச் சாதனங்களை இரவல் பெறுவது போன்ற செயல் முறை மற்றும் சேவையை

மேம்படுத்த உதவுவது.

இ) நூலக சாதனங்கள் பற்றிய தகவல்களைப் பயனீட்டாளருக்கு எளிதில்

தெரியப்படுத்துவது.

ஈ) பயனீட்டாளர் இரவல் பெறுவதற்கு விரும்பும் பொருளை விரைவில் கண்டறிவது.

உ) காகிதப் பயன்பாட்டைக் குறைக்கும் சூழலை உருவாக்குவது.

மலாயாப் பல்கலைக்கழக மின்னியல் நூலகச் சேவைகள்

தற்போது பல்வேறு வகை மின்னியல் நூலகம் பயன்பாட்டில் உள்ளன. மலாயாப் பல்கலைக்கழகத்தைப் பொருத்தமட்டில், மலாயாப் பல்கலைக்கழக செய்திகள், மலாயாப் பல்கலைக்கழகத்தின் முந்தைய ஆண்டுகளின் தேர்வுத் தாள்கள், மலாயாப் பல்கலைக்கழக நினைவு மலர்கள், மலாயாப் பல்கலைக்கழக மின் அச்ச ஆகியவை மின்னியல் முறையில் தயாரிக்கப்பட்டு பாதுகாக்கப்படுகின்றது. மலாயாப் பல்கலைக்கழக மின்னியல் நூலகம் வழங்கும் சேவைகள், உட்பொருள் வகைகள் குறித்த விரிவாக்கம் கீழே வழங்கப்படுகிறது.

- கல்வி மையங்கள் சார்ந்த தகவல்கள்
- நூல் குறிப்பு, நூல் முகப்பு மற்றும் பொருளடக்கம், பதிவுக் குறிப்பு, நூல் ஆய்வு மற்றும் ஆலோசனைகள், நூல் விற்பனைத் தகவல்
- மின்-நூல், மின்-சஞ்சிகைகள், ஆய்வுக் கட்டுரைகள்
- அறிஞர் குறித்த தகவல்கள்
- பதிவுகள், கலாச்சார உட்பொருள்கள், அறிவியல் தொழில் நுட்பம், தரம், பதிப்புரிமை
- பொதுமக்களின் அரசியல் தகவல்
- அரசாங்க அலுவலகங்கள் மற்றும் பொது நிறுவனங்கள் வழங்கிய மின்னியல் உட்பொருட்கள்
- பொதுமக்களின் அரசியல் பதிவுகள், ஆண்டறிக்கை, கணக்காய்வு மற்றும் இதர உட்பொருட்கள்
- வெளி நாடுகளின் ஆவணங்கள்
- உள்நாட்டுத் தகவல்
- உள்நாட்டு அரசாங்க அலுவலகங்கள் மற்றும் உள்நாட்டுப் பொது நிறுவனங்கள் வழங்கிய மின்னியல் உட்பொருட்கள் (நூலகங்கள், பொருட்காட்சி சாலைகள், கலாச்சார அமைப்புகள், மேலும் பல)
- உள்நாட்டு ஆய்வு அறிக்கை, ஆண்டறிக்கை, கணக்காய்வு, உள்நாட்டுக் கொள்கை ஆவணங்கள்
- உள்நாட்டு இடங்கள், செய்திகள், கலாச்சாரம், மேலும் பல தொடர்புடையத் தகவல்கள்
- வெளிநாட்டுத் தகவல்

- வெளிநாட்டு கல்வி சார்ந்த மாநாடுகள், கல்வி சார்ந்த ஆவணங்கள், மேலும் பல
- பல்வேறு கலாச்சார செய்திகள், அன்றாடத் தகவல், கலாச்சார தகவல், பல்வேறு கலாச்சார சிக்கல்கள் தொடர்பான கொள்கைத் தகவல்.
- பல்வேறு கலாச்சார சிக்கல்கள் - செய்திகள் கொண்ட அகப்பக்கங்கள்
- சட்ட தகவல், கணக்காய்வு, அங்கவீனர்கள் குறித்த ஆய்வுப் பதிவுகள்
- அங்கவீனர்கள் குறித்த செய்திகள் - சார்ந்த சிக்கல்கள், சிகிச்சை மையங்களின் தகவல், அன்றாடத் தகவல், கல்வித் தகவல், போக்குவரத்துத் தகவல்
- அங்கவீனர்களுக்கான அகப்பக்கங்கள்

ஏறக்குறைய பழைய பயன்பாட்டு முறையில் இருந்த பெரும்பான்மையான நூலகச் சாதனங்களையும் மின்னியல் முறைக்கு மாற்றம் செய்தாகிவிட்டதோடு அது குறித்த தகவல்களை அகப்பக்கத்திலும் வழங்கப்பட்டுள்ளது.

மலாயாப் பல்கலைக்கழகத்தை முழுமையான மின்னியல் முறைக்குக் கொண்டு செல்ல ஆலோசனைகளும் ஆயத்தங்களும்

மலாயாப் பல்கலைக்கழகம் நூலகச் சாதனங்களைக் கையாளுதலை, எளிய அணுகுமுறையும் சிறந்த செயல்பாடும் கொண்ட வகையில் மேலும் மேம்படுத்த எண்ணம் கொண்டுள்ளது. முழுமையான மின்னியல் நூலகத்தை நோக்கி மலாயாப் பல்கலைக்கழகம் செல்வதற்கான ஆலோசனைகளுள் சில:

அ) உறுப்பினர்களுக்கான மின் பதிவு முறையை உருவாக்குதல்

ஆ) நூலகச் சாதனங்களின் வரிசைப்படுத்தப்பட்ட பட்டியலும் இரவல் வாங்கக்கூடிய தரவுகளின் பட்டியலும் கிடைக்குமாறு செய்தல். இதன் வழி சேவைத் தரத்தை உயர்த்தவும் நூலகச் சாதனங்கள் குறித்த தகவல்களை எளிதாக, விரைவில் அறிந்து கொள்ளவும் முடியும்.

இ) பலதரப்பட்ட உலகளாவிய தகவல்களை வழங்கும் இடமாக மலாயாப் பல்கலைக்கழக அகப்பக்கத்தின் சேவையை மேம்படுத்த வேண்டும்.

ஈ) மின்னியல் நூலகத்தில் நூல், ஆய்வுக் கட்டுரைகள் மற்றும் இதர தகவல் சாதனங்களைப் படிப்படியாகச் சேர்த்து, அவற்றைக் கண்டறிய தக்க குறியீட்டுப் பட்டியலை அவ்வப்போது புதுப்பிக்க வேண்டும்.

உ) தற்போதைய நூலக முறையைச் சிறந்த கூடுதல் சேவைகள் வழங்கக் கூடிய புதிய முறைக்கு மேம்படுத்த வேண்டும்.

மலாயாப் பல்கலைக்கழகத்தை முழுமையான மின்னியல் முறைக்குக் கொண்டு செல்வதில் ஏற்படும் சிக்கல்களும் சவால்களும்

மலாயாப் பல்கலைக்கழகத்தில் முழுமையான மின்னியல் நூலகம் ஏற்படுத்தப்படும்போது நிறைய சிக்கல்களையும் சவால்களையும் மலாயாப் பல்கலைக்கழக நூலகம் சந்திக்க வேண்டியிருக்கிறது. குறிப்பிட்டுச் சொல்லக்கூடிய சவால்களும் சிக்கல்களும்:

- நூல்கள், கருத்தரங்கு மற்றும் நடவடிக்கை ஆவணங்கள், பதிவுப் பேழை, புகைப்படம் போன்ற நூலகச் சாதனங்கள் பல இன்னும் பதிவேற்றம் செய்யப்படாத நிலையில் உள்ளன. இவை பதிவேற்றம்

செய்யப்படுவதற்கு முன், நல்ல நிலையில் இருக்கும் பொருட்டு புதுபிக்கப்பட வேண்டும். நூலகப் பொருட்களைப் புதுப்பித்தல் மற்றும் பராமரிப்பு வேலைகள் முதலில் நடப்பதால் அவற்றுக்கு ஆகும் செலவு பதிவேற்றத்திற்குத் தடையாக அமையும்.

- அதோடு, பதிவேற்றப்பட வேண்டிய பொருட்கள் குறிப்பிட்ட தரத்தில் இருப்பதாலும், அவற்றை நூலகத்தை விட்டு வெளியே கொண்டு செல்ல முடியாது என்பதாலும், பதிவேற்றத்திற்குத் தேவையான இடம், அடித்தளம் அமைத்தல், பதிவேற்றக் கருவி, சேமிப்புக் கருவி போன்றவற்றுக்காகப் பதிவேற்றப் பணிக்கு அதிகப்படியான தொகை தேவைப்படும்.
- பொருட்களைப் பராமரித்தல் மற்றும் புதுப்பித்தல் பணிகள் குறிப்பிட்ட தரத்தில் செய்யப்பட வேண்டும் என்பதால், மலாயாப் பல்கலைக்கழகத்திற்கு அந்தப் பணிகளுக்கான சிறந்த வல்லுனர்கள் தேவை. அதனால், மலாயாப் பல்கலைக்கழகம் நூலகம் தொடர்பான அரசாங்க நிறுவனங்களான தேசிய நூலகம் மற்றும் மலேசிய பழஞ்சுவடிக் காப்பகம் போன்றவற்றின் கருத்தையும் உதவியையும் பெற வேண்டும்.
- இந்த மின்னியல் நூலகத்தின் நடைமுறையின் தரத்தை உயர்த்த, மலாயாப் பல்கலைக்கழகத் தரப்பினர் புதிய சூழலில் தங்களைப் பொருத்திக் கொள்ளத் தேவையான பயிற்சி, பட்டறை, கருத்தரங்கு போன்றவற்றைப் பெற தொழிலாளர்களுக்கு வாய்ப்பளிக்க வேண்டும். இதன்வழி தொழிலாளர்கள் பெறும் அறிவுத் திறன் மலாயாப் பல்கலைக்கழகத்தைச் சிறந்த முறையில் கொண்டு செல்லவும் பல்வேறு பயனீட்டாளர் தேவைகளை முறையாகக் கையாளவும் வழிவகுக்கும்.
- மலாயாப் பல்கலைக்கழகம் போதிய பாதுகாப்பு முறையை மேம்படுத்த வேண்டும். நூலகச் சாதனங்களைப் பராமரித்தல், புதுப்பித்தல் மற்றும் மின்னியல் பதிவேற்றம் செய்தல் ஆகிய பணிகளில் தேவையற்றோருக்குத் தகவல்கள் கசிய வாய்ப்பு உள்ளது. அதனால், பாதுகாப்புக் குழு ஒன்றை ஏற்படுத்தி, தகவல் மற்றும் மின்னியல் நூலக முறைமையையும் பயனீட்டாளர்கள் கருத்தையும் பாதுகாக்கச் செய்யலாம்.

சிக்கலுக்கான தீர்வு வழிகள்

மேற்கூறிய சிக்கல்களையும் சவால்களையும் தீர்க்க ஒரு சில வழிமுறைகள்:

- மின்னியல் நூலகம் சிறந்த தனித்துவம் வாய்ந்த தகவல் மையமாக இருக்க மலாயாப் பல்கலைக்கழகம் பல தரப்பினரின் ஆதரவையும் ஒத்துழைப்பையும் வல்லுனர்களின் சேவைகளையும் பெற வேண்டும்.
- நூலகப் பொருட்களைச் சிறந்த வழியில் பாதுகாத்து, சீரிய அமைப்பில் இருக்க பராமரிப்பு, புதுப்பித்தல் மற்றும் பதிவேற்றம் போன்ற பணிகளுக்குப் போதிய தொகை வழங்கப்பட வேண்டும். அதன்வழி மின்னியல் நூலகக் கனவை முழுமையாக நனவாக்க முடியும்.
- மின்னியல் நூலகத் திட்டத்தில், மலாயாப் பல்கலைக்கழகம் மின்னியல்படுத்த வேண்டிய நூலகப் பொருட்களைக் கண்டறிவதோடு, இத்திட்டத்திற்கான நோக்கம், அடைவு, இலக்கு ஆகியவற்றிற்கு ஆய்வு மேற்கொள்ள வேண்டும். அதோடு மலாயாப் பல்கலைக்கழகம் ஒவ்வொரு துறையிலும் உள்ள தொழில்நுட்பச் செயற்குழுவோடு கலந்தாலோசித்து மின்னியல் நூலகத்தில் சேர்க்க விரும்பும் தகவல் சாதனங்களுக்கான தேர்வுக் குறிப்பு ஏற்படுத்த வேண்டும்.
- பாதுகாப்புப் பகுதியில், பராமரிப்பு, புதுப்பித்தல் மற்றும் பதிவேற்றப் பணிகள் மலாயாப் பல்கலைக்கழக வளாகத்தில் தக்க அதிகாரிகளின் மேற்பார்வையில் நடைபெற வேண்டும். இதன்வழி நூலகப் பொருட்கள் கட்டுப்படுத்தப்பட்டவையாக இருந்தால் அதன் தகவலை பிறர் அடைவதில் இருந்து காக்க முடியும்.

- நூலக ஊழியர்களின் ஒத்துழைப்பினாலும், தகவல் தொழில்நுட்பம் மற்றும் தொடர்பில் இருக்கும் அறிவுத் திறனாலும், மின்னியல் நூலகப் பணி சீரிய முறையில் நடக்கும். இருப்பினும், மின்னியல் தகவல் கையாளுதலுக்கு, நூலக மற்றும் மேம்பாட்டுப் பகுதி ஊழியர்களுக்கு முழுமையான பயிற்சி வழங்குதலை கட்டாயமாக்க வேண்டும்.

முடிவுரை

தகவல்களைப் பயனீட்டாளர்களுக்கு வழங்கும் காப்பகமாகவே நூலகங்கள் எல்லாக் காலங்களிலும் செயல்பட்டு வந்திருக்கின்றன. அதனை அடுத்த பரிணாமத்திற்குக் கொண்டு செல்வதே இந்த மின்னியல் நூலகத் திட்டத்தின் அடிப்படைத் திட்டமாகும். இதற்காகவே மலாயாப் பல்கலைக்கழக நூலகமும் இத்திட்டத்திற்கு அதிக அக்கறை காட்டி செயல்பட்டு வருகின்றது. ஆராய்ச்சி மற்றும் புதுமைகளை அவ்வப்போது அரங்கேற்றி வரும் உலகளாவிய கண்டுபிடிப்புகளுக்கு ஈடுகொடுத்து, ஆய்வாளர்களுக்குப் போதிய தரவுகளை வழங்குவதில் நூலகங்களுக்கு மிகப் பெரிய கடமை உள்ளதை உணர்ந்து, பிரச்சனைகள் பல இருப்பினும் அவற்றைக் களைந்து, உலகளாவிய மாற்றங்களுக்கு ஈடு கொடுத்து முழுமையான மின்னியல் நூலகத் தரத்தை அடைய முன்னேறுவதே மலாயாப் பல்கலைக்கழக நூலகத்தின் குறிக்கோலாகும்.

Development of Tamil Digital Libraries: Advances and Challenges

Dr. K. Kalyanasundaram
Lausanne, Switzerland

Introduction

Computers and Internet possibly are the two most important technological innovations of the last century that had a pronounced impact on the humanity. Together they have changed dramatically the way people store information and exchange them with others. On the educational front, the impact in the Indian sub-continent is more pronounced. Ways and means by which information (knowledge?) has been stored and transferred from generation to generation has been evolving over many centuries: from “gurukulam” system [1] (direct largely vocal transmission of knowledge from the teacher to the student) to inscriptions in caves, copper plates to written texts as in palm-leaf manuscripts to printed books. The contents of 26-volume summary - a systematic collection of inscriptions of south India compiled by Prof. Hultzsch, published by the Archeological Survey of India is available online [2]. At the dawn of the 21st C, information storage and exchange is moving to electronic form. Net-based information interchange available widely and at low cost has shrunk the distance barriers that exist amongst friends and relatives living in far off places. Growing number of educational tools are produced in electronic form as e-books, multimedia-based teaching software and even distance/remote education through online portals. In this paper we focus on one important component in distance education and information sharing through the net, viz., digital libraries with exclusive focus for Tamil language.

For any society or community, its cultural heritage is measured by the variety and level of content available in key areas such as literature and performing arts (music, dance and drama). Tamil is one of the oldest and living languages of the Indian sub-continent, with a vibrant history that dates back at least to 2000 years. Tamils have a remarkable cultural heritage, evidenced by huge and rich repertoire in all the key areas mentioned above. Preservation and propagation of this rich cultural heritage in a form compatible with contemporary trend and usage mode is essential. For many reasons, Tamil heritage information is largely preserved in forms that are at great risk. There are thousands of works in science and literature still in palm leaf manuscripts, not yet published in printed book form. With poor storage conditions, existing collection of palm leaf manuscripts and printed books is degrading rapidly. In the case of Sri Lanka, huge and precious collections of literary works of Tamils for several centuries simply vanished when the Jaffna Library got burnt in a major fire incident [3]. Anthropologists are concerned about rapid changes in the life-style of Tamils with a rapid decline in the practice of rich “folklore arts” that are unique. Even in areas such as archeology, Tamils are proud to have a vast collection of temples of varying sizes and artistic features, each with its own unique mural paintings and other treasures [4]. Sadly these are also not being preserved in a form that can guarantee posterity. In view of this Tamil Diaspora have an important moral obligation to start e-preservation efforts in the areas outlined above.

Objectives of digital libraries

It is useful to start with the simple question, “what is a digital library”. Widely accepted definition of a digital library is one of “a comprehensive « networked » information environment, a seamless rich set of tools and resources for the user community, accessible round the clock across the globe with the power of the Net. Digital resources can be in different formats and for different purposes, different target audience(s), accessible online or offline and derived from various « primary, existing » resources in different formats (books, audio/video recordings, manuscripts and photographs). Of late many information sources are produced directly in digital or e-form.

Digital Catalogue of Tamil Materials and Digital Dictionaries

A first step in any digitization effort is assessment (stock-taking) of diverse form of cultural heritage materials that are still available, a catalogue of “who has what and where?” While a catalogue of the library holdings in digital form “searchable” remotely is routine and exists for nearly all the libraries in the west, the same cannot be said about major libraries of Tamilnadu. There is no central inventory of public and private collections of cultural heritage related materials of Tamils within and outside India. Tamil Diaspora has migrated in large numbers to several far-off places (e.g., Sri Lanka, Malaysia, Singapore, Fiji and Mauritius to name the major ones) and developed their own local “identity” over the years.

Academic Institutions of North America and Western Europe have been working together to build “union catalogue”, a central inventory of “who has what and where?” Many countries are building “union catalogue” of regional collections. Two notable ones are i) CRL (Center for Research Libraries, an international consortium of university, college, and independent research libraries [5]; ii) “Worldcat”, sponsored and run by OCLC (Online Computer Library) a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing the rate of rise of library costs [6]. More than 72,000 libraries in 171 countries and territories around the world use OCLC services. Nearly 1.5 billion items in all world languages held in different libraries worldwide are catalogued here. A notable feature of the Worldcat is that their database can be searched directly in Tamil script in Unicode.

For languages of south India, Digital South Asia Library (DSAL) is a major initiative of CRL managed by the University of Chicago [7]. An important component of the DSAL is their *Digital Dictionaries of South Asia* [8]. In addition to the classic Madras University Tamil Lexicon, digital searchable versions of important Tamil dictionaries of Fabricius, Kadirvelu_Pillai, McAlpin and Winslow are available online. The *South Asia Union Catalogue*, initiated by CRC and managed by DSAL, intends to become an historical bibliography comprehensively describing books and periodicals published in South Asia from 1556 through the present [9]. In addition, it will become a union catalog in which libraries throughout the world owning copies of those imprints may register their holdings. For an overview of some of the digital south asia library efforts, see the special issue of Focus, vol. 24, number 3 (spring 2005), PDF downloadable from CRL website [10]. Roja Muthiah Research Library based in Chennai is an important and modern library for Tamils with over 100,000 volumes of books, journals, and newspapers in their shelves [11]. A digital catalogue of RMRL library is available via the DSAL gateway.

INFLIBNET is a Govt. of India UGC supported consortium network and is slowly evolving as the major information gateway on the holdings of Indian Universities [12]. Vidyanidhi is another Digital library initiative of India to facilitate the creation, archiving and accessing of doctoral theses [13]. Vidyanidhi is envisioned to evolve as a national repository and a consortium for e-theses through participation and partnership with universities, academic institutions and other stakeholders. "Indcat" project of INFLIBNET is unified Online Library Catalogues of books, theses and journals available in major university libraries in India [14]. Over 11 million books and twenty thousand doctoral theses are included in the "indcat" database.

Digital Preservation of Tamil Literary works

One of the earliest preservation modes for printed books, still widely in use, is to take microfilm copies of the work. Three main advantages of preservation as microfilm are compact size, long bench life of microfilms (at least 75 years) and facile browsing the content using a microfilm reader. Modern version of the microfilm readers permits even digital capturing of pages of the microfilm or print out the page. University libraries of major institutions in north America and Europe still prefer microfilm form of preservation. Roja Muthiah Research library of Chennai has been making microfilm editions of Tamil works for more than a decade – possibly the only organized effort for Tamil literary works.

The second popular method of digital preservation of works is in the form of image files scanned at high resolution. The main advantage of this approach is that the digital image preserves all the presentation details (layout, artistic drawings, calligraphy used to present the work). Disadvantages are that the file size are huge and the content of the image files not easily searchable. Digital Library of India initiative, launched as part of a larger "million ebooks" project of the Carnegie-Mellon University focused on the digitization of works in various languages of India. Image files in the form of "tiff" format have been made for several thousand works. Tamil books available in e-form at DLI is only few percent, many duplicate files of the same work scanned at different scanning centers.

Recent initiative of Google to digitize collections of major university libraries of US falls in the same category – ebooks in the form of scanned image files [15]. Interestingly for many European languages, good OCR (optical character recognition) software's exist and they can be used on the scanned image files of printed books to generate a machine-readable version. Google's online ebooks interface for browsing the content of their eBook collections in fact has enabled this OCR processing to export the equivalent text of a text page displayed as an image file. Unfortunately for Tamil we do not have high performing OCR software for direct use on the image file version of eBooks.

In the last decade several initiatives have been launched for digital preservation of Tamil works. Mention here must be made of Tamil Heritage Foundation [16], Noolaham.net [17] and that of Pollacchi Nesan [18]. All of them produce etexts as image files in tiff or jpg format. Noolaham focuses on the works (books and newspapers) of Tamil authors of Sri Lankan origin. An urgent need in the digital preservation exercise of Tamil Heritage related materials is a central database (union catalogue) of what has been generated as of date. Without a union catalogue, precious human and machine resources are being wasted in digitization of the same work/object by several initiatives.

Machine Readable Version of etexts

There have been several projects to prepare machine-readable version of ancient works (those in public domain) for free distribution on the Net. Pioneering effort has been Project Gutenberg [19], where over 30000 free eBooks are made available to read on the PC and in a number of portable devices such as iPhone and dedicated ebook readers [20] such as Kindle and Sony eBook Reader. Dr. Thomas Malten of the Institute of Indology and Tamil Studies of the Univ. of Koeln, Germany possibly was the first person to take initiative to digitize Tamil works in machine-readable form in early nineties [21]. Supported by grants from the German Government, he prepared electronic texts in Romanized (transliterated) format of Tamil works of first sangam period and few major works such as Kamba Ramayanam. The etexts used a simple plain ASCII based transliteration scheme so that the etexts can be readily searched. An online search interface was made available to search the entire collections free of cost, an extraordinary feat at a time where the computer usage by the linguists and Tamil Diaspora was not wide, as it happened after the arrival of low cost personal computers. In the early nineties, free Tamil fonts started appearing and also software such as Adami that permitted text input in Romanized format with the option to display the equivalent text natively in Tamil script form. Parthasarathy Dileepan of Tennessee US led a maiden initiative to produce e-version of the entire 4000 verses of nalayira divya pirapandam. Using Adhavin (windows version of Adami) software, it was now possible to view and print the etext verses in Tamil script form. Then came Project Madurai devoted to preparation of etexts directly in Tamil script format. Tamil Virtual University, launched in 2001 has included a digital library section which carries etexts of several Tamil works as web-pages [22].

Project Madurai

Project Madurai (PM) was launched in 1998 as a voluntary Net-based initiative for the digital preservation and propagation of Tamil literary works through Internet, natively in Tamil script form [23]. Volunteers from all four corners of the world use their spare time and personal computer equipment to prepare etexts of Tamil literary works natively in Tamil script form. During the past ten years, over 400 Tamil works, big and small, have been digitized and distributed through a dedicated web-server. The etexts are distributed FREE as formatted text in HTML and PDF formats. Anyone can download for personal use and forward to others. Only requirement for reproduction is that the credit acknowledgement lines included in the header part of the etext be preserved.

The coverage is very broad in scope, spanning a wide time period and include works from early Sangam period to contemporary literature: pattuppATTu, eTTuttokai, patinenkIzkaNakku, epics, religious works of saivaite, vaishnavite canon, Old and new Testament of Bible in Tamil, ciRappuRaNam, saiva siddhanatha works, works of Bharathi, Bharathidasan and Kalki and works of Sri Lankan and Malaysian Tamil authors to mention broad categories. Only condition is that the covered Tamil work should be free of copyrights (work of “public domain”) or the author, their legal heirs give permission for royalty-free reproduction of the work and free distribution of the etext worldwide. The coverage includes all “genre” – poetry, prose and drama, English translations of important Tamil works as well Tamil translation of key literature of other world languages.

Challenges faced by PM during the past decade

Project Madurai is the only initiative devoted to preparing machine-readable (searchable) form of etexts. Nearly all other initiatives of digitization store the etexts in the form of image files in tiff, jpg and other formats. For web-delivery of Tamil etexts in a form usable by any average user of Tamil worldwide, several conditions are to be met. For machine-readable texts, important question is the font encoding to be used. Unlike most of the languages of the west, for Tamil, there has not been universal agreement on the font encoding to be used for electronic version of digital data. If the encoding used is not a standard for use in all computer platforms, then suitable Tamil fonts that work flawlessly in all platforms are to be made available free and the font encoding used should be such that there is no corruption of the data during transfer across platforms. When Project Madurai was launched in 1998, we had only two Tamil fonts (Inaimathi and Mylai) were available that satisfied above requirements. Hence etexts were prepared in these two font encodings. Soon there was a Net-based encoding TSCII (Tamil script code for Information Processing) evolved as a popular encoding for Tamil and Project Madurai started releasing etexts in this 8-bit bilingual encoding. It may worth pointing out here that flawless delivery and processing of Tamil digital materials require use a bilingual 8-bit scheme with standard ASCII scheme as part of the scheme. Around 2000 windows PC started supporting Tamil in the multilingual Unicode encoding scheme. As a consequence, Project Madurai started distributing etexts in Unicode format as well. As of date, PM etext collections are available in two encodings TSCII and Unicode.

Project Madurai has been fortunate to have a steady team (small, max 20. at a time but fully committed) of volunteers who contribute hours in the preparation of the etexts. They are Tamil enthusiasts keen to see our literary heritage preserved in e-form but are based in various far-off places without access to a good collection of printed copies of the Tami literary works. Persons who had time to key-in the work or to proof-read the work typed by other (PM works are proof-read at least once independently by a second volunteer) do not have access to printed books and those who have a good personal collection do not have time to contribute to the project. As Project Managers we have to go periodically to different bookshops in Chennai to procure target books for digitization. Sadly Tamil books covering good part of Grammar, prose and poetry are no longer reprinted due to pronounced shift in the interests of Tamils towards novels in late 20th C. So we are obliged to depend on the precious collections still preserved in major public libraries of Tamilnadu. Janet Library of the Univ. of Koeln, Germany is unique in this context to host possibly the biggest collection of Tamil books outside India, with collections over 60,000. Unfortunately financial resources available to the Indology Institute IITS there (where Prof. U Niklas and Dr. T Malten are associated) are extremely limited and the Institute even had passed through total shut down few years ago.

In order to improve the accessibility of books to our volunteers willing to type in the text, PM adopted the “distributed proof-reading (DP)” approach developed by Project Gutenberg. In DP, scanned image files of printed books are collected and stored in a web-server. Volunteers willing to participate in the eBook preparation access these image files, one page at a time, either for key-in or proof-reading. Using special software, the image of a printed page is displayed on the left in a split-screen window with a text-editor on the right to key in (or proof-read) equivalent text. Using this distributed proof-reading implementation at PM (DP-PM) [24] we have been able to produce nearly 100 ebooks during the past few

years wherein a single eBook was produced as a joint effort by a group of volunteers based in different parts of the globe.

Digital Library of India initiative [25] has been very useful to us in supplying image files of several Tamil literary works and we are very grateful to them for this wonderful effort. Unfortunately funding for the DLI project has ceased, though the collections are still available on the Net. One serious limitation of the DLI collections is that the integrity of the text cannot be guaranteed. The image files have been not systematically checked with the source after editing of the graphic files. For several works often few lines at the top or bottom of the page are missing or words clipped on the right or excessive cleaning of the contrast resulting in pure consonants and akara-varisai abugida characters appear in the same manner (as is the case with palm-leaf manuscripts). With such pitfalls, we are obliged to consult the original printed version to ensure textual accuracy of the etext being produced.

Another important factor to consider is the selection of an “authentic” edition of a given work for the etext preparation. Some works have ancient works reproduced without much of sandi/word splitting so as to make the work more understandable to the lay public. For many ancient works there are textual variants (pATa pEdam). One key decisive factor at least for works of poetry in nature is the compliance to the grammar rules (metrics). Checking of a given work for metric accuracy requires a higher level of linguistic expertise, not often available with the volunteers who help us produce the etexts. In this context it is preferable that the entire collection of PM etext collections are vetted by a team of language experts, possibly by a team of university level researchers. Tamil scholars like UVeCa compared several editions of ancient Tamil works to compile “authentic” or “critical” editions.

Project Madurai etext collections at best are raw texts as they were written. In his presentation at the last year’s Tamil Internet 2009 conference, Dr. Jean-Luc Chevillard proposed creation of a second generation of etexts for scholarly research [26]. Suggestion was to produce “critical” editions where individual words of verses are “tagged” with indication of the ciir for each word of the verse. Dr. Chevillard pointed out “Text Encoding Initiative (TEI)” provides a number of modules which can be applied (or adapted) to the Tamil case. An etext can be “metric compliant” or “sandi-split”. Since the meaning of verse can vary with the words, contemporary “sandi-split” versions are in reality an interpretation of a literary work. In principle, with the knowledge of the “sandi-splitting” rules and information on the metrics of a given work, sandi-split versions can be converted to metric compliant versions with the use of a software. In fact In the last TIC 2009 conference, Balasundararaman presented a model software that permit checking of poetic works for their metric compliance (venba metrics) [27]. Project Madurai just started extending the coverage of Tamil works with the inclusion of commentaries of works. There are “classical” commentaries by scholars such as parimelazhakar, naccinarkiniyar that must be preserved in digital form.

The advantage of a machine-readable version of an etext is that all the words of the work can be indexed in a database in several ways, permitting facile search on the occurrence of a given word or a string of words (word sequence) in one or more literary works. With a vast collection of etexts of Project Madurai covering Tamil works spanning a wide time frame over two thousand years, such database-driven search for specific words can be very useful in etymology studies by the language experts. We have provided prototypes of such searchable online interfaces for Tamil works, using databases based on MySQL and

php query calls. Dr. Vasu Renganathan of the Univ. of Pennsylvania for example, has undertaken such systematic studies of etymology using etext collections of Tamil literary works.

The digital edition of “tEvAram” published by the Pondicherry Institute of Indology is an excellent example of combining multimedia tools with machine-readable text to bring an ebook of the next generation [28]. For each tEvAram verse, in addition to the text, audio version of the verse and graphic (map showing the location of the town where the temple/deity referenced in the text is given) files are added to enhance the utility of the electronic version. Online websites for the teaching of Tamil of the Tamil Virtual University and of Univ. of Pennsylvania Tamil web have nicely integrated multimedia tools to the electronic text of the Tamil work taken up for detailed study.

Concluding Remarks

There are areas such as computer-aided teaching of Tamil online or off-line need a comprehensive digital library. In pioneering efforts, Tamil Virtual University and Tamil Web of the University of Pennsylvania have put together a number of educational tools – lessons as web pages with a number of audio- and video clips integrated. The Ministry of Education of the Singaporean Government also has been supporting development of numerous multimedia tools to aid teaching of Tamil at the primary and secondary school level. Throughout this article we have indicated numerous public-funded and voluntary efforts working in the area of Tamil digitization. An umbrella organization that can interface all these isolated efforts can go a long way in reducing redundancies and accelerate output. In addition to the governmental agencies, academics and IT/ICT professionals have important role to play in this key area of academic and public interest.

Bibliography:

1. <http://www.lifesciencefoundation.org.in/page6.html>
2. <http://www.whatisindia.com/inscriptions/>
3. http://en.wikipedia.org/wiki/Burning_of_Jaffna_library
4. http://www.tamilartsacademy.com/articles/list_of_articles.html
5. <http://catalog.crl.edu/>
6. <http://www.worldcat.org>
7. <http://dsal.uchicago.edu/>
9. i) <http://www.crl.edu/focus/article/509>; ii) <http://sauc.uchicago.edu/>
10. <http://www.crl.edu/focus/spring-2005>
11. <http://www.lib.uchicago.edu/e/su/southasia/about-rmrl.html>
12. <http://www.inflibnet.ac.in/>
13. <http://www.vidyanidhi.org.in/>
14. <http://indcat.inflibnet.ac.in/indcat/>
16. <http://www.tamilheritage.org/>
17. <http://www.noolaham.net/>
18. <http://www.thamizham.net>
20. <http://ebook-reader-review.toptenreviews.com/>
21. i) <http://www.uni-koeln.de/phil-fak/indologie2/> ii) <http://webapps.uni-koeln.de/tamil/>
22. <http://www.tamilvu.org>
23. <http://www.projectmadurai.org>
25. <http://www.new.dli.ernet.in>
26. <http://www.linguist.univ-paris-diderot.fr/~chevilla/>
27. <http://www.infitt.org/ti2010/>
28. <http://www.ifpindia.org/Tamil-Saiva-Hymns.html>

Reducing Digital Divide in Tamilnadu using Data Mining Techniques for better E-Governance

R. Jayabrabu
M.C.A.,M.Phil.,(Ph.D.,)
Assistant Professor
Department of Computer
Application
School of Science and Humanities
Karunya University
Coimbatore - 641 114
Tamilnadu
E-mail: jayabrabu@rediffmail.co.in

Prof. (Dr). V. Saravanan
M.C.A.,M.Phil.,Ph.D.,
Professor & Director
Department of Computer
Applications
Dr. N.G.P Institute of
Technology
Coimbatore - 641 048,
Tamilnadu
E-mail: profsaran@hotmail.com

Er. A.K. Balakrishnan
M & B Associates
75 Naichimuthu Gounder
Colony
Sanganoor Road
Ganapathy
Coimbatore - 641 006,
Tamilnadu
E-mail:
cbebalakrishnan@gmail.com

Abstract

The term digital divide refers to the gap between people with effective access to digital and information technology and those with very limited access. In other words it is closely related to the knowledge divide or knowledge share due to the lack of technology and knowledge. The extraction of useful and non-trivial information from the huge amount of data available in many and diverse fields of science, business and engineering is called as Data Mining. Data Mining techniques and algorithms are the actual tools that analysts have at their disposal to find unknown patterns and correlation in the data. For effective use of E-governance in Tamil Nadu, the digital divide to be reduced. Most of the Government departments are already using E-governance in Tamil Nadu. This is the appropriate time for us to analyze the effectiveness and reach ability of technology to all sectors of peoples. Even the most learned peoples are reluctant in using the technology. This digital divide gap leads to improper usage of Information and Communication technologies. The objective of this paper is to analyze the following two important digital divide issues using data mining and present recommendations for better E-Governance in Tamil Nadu.

a. Improving Quality of Bandwidth/Parameters.

Since, the information and communication technologies are being implemented in the Government at different levels; good bandwidth is needed for constant transformation of knowledge in a proper format. For the better usage of E- Governance, the quality and performance of bandwidth performance has to be increased. In Tamil Nadu, there are so many service providers available for connectivity. But, the expected quality of bandwidth is less than the assured bandwidth. This paper analyses the bandwidth parameters using data mining techniques and suggest a better framework for improving bandwidth across Tamil Nadu.

b. Taking Technology to reduce the gap

Now a day, many new information and communication technologies are introduced. The urban sector is reluctant in using the technology due to the fear of using the technology and also thinking communication/network failure, which occurs frequently. The middle age person still thinks that the technology is very far from them and also is very costlier. The rural sector is unaware of these technologies and they need to be provided infrastructure and training. With the increase usage of mobile phones; convergence of technologies also need to be thought of. This paper analyses the need of urban and rural sector people for the effective reach of technology. Data Mining Techniques are used for data analysis, which leads creation of to better E-governance standards.

The above mentioned parameters are studied by applying data mining techniques such as Association rule mining (determine implication rules for a subset of record attributes, Classification (assign each record of a database to one of a predefined set of classes analysis and Clustering Techniques (find groups of records that are close according to some user defined metrics) and a suitable framework is proposed for better E-Governance.

1. Introduction

When the IT industry increased globally in the 19th century, simultaneously the Internet and the Mobile technologies are also emerged into the world and ruled majority of the people. With this, E-Governance also booms out with the help of some Government Departments around the world. In India, National Informatics Center (NIC) played a vital role for the development of E-Governance in which they incorporate some of the Government related activated like Tax payment, Census Generation, Election Management, Disaster Management,[1] etc.,

In Tamil Nadu, some of the successful E-Governance projects are land registration, call for tender, issue of birth/death certificates, agriculture, e-transaction, RTO, tourism, infrastructure, land/local tax, local body election details, e-ticket etc., [1]. The major scenario in the above-mentioned successful E-Governance is heterogeneous based System. The entire activities of each Government related activities posses a unique database to store their respective data. This technique is followed in our state and also other states too [2]. As a fact, each Government department maintains their own database as unique and there is no interlinking between various departments/databases. When, the land registration department needs some information about agriculture data, they are not able to access the agriculture database. This leads to minimum usage of the e-governance projects by the citizens. The digital gap increases due to the issue and the important e-governance projects fails after implementation. This paper proposes the use of data mining techniques and a better framework to reduce the digital gap and to interlink the heterogeneous databases. This paper proposed two stages to reduce the digital gap

- a. Improving Quality of Bandwidth/Parameters.
- b. Taking Technology to reduce the gap

2. Using Data Mining Techniques to Reduce the Digital Gap

Data Mining is the technique to explore and analyze the large data sets, in order to discover meaningful patterns and rules [6]. The evaluation of data mining techniques began when the business data are stored

in the database and the technologies were generated to allow the user to navigate the data in the real time.

Recently, the ICT made a proposal for all the state and central Government for the betterment of database maintenance in the near future generation [3]. As we know, now a days, all the Government departments utilizes huge amount of data in their day-to-day work, which leads to maximize the access of current or history of datasets from the database [2]. But, it is not possible to fetch the datasets when they need. This is because of insufficient data, improper format, duplicated data, and some technical problems etc.;; When we discuss on other side, it is also due to less bandwidth, natural disaster, network failure, and loss of data during data transmission and collision of packet with one another etc. As a result of this, the end user cannot able to perform the operation with in the time and also little afraid to continue the E-governance system. Since, a gap is generated between user and existing E-Governance systems. As a result, the Government should concentrate on above set problems for the betterment of E-Governance. By considering these issues, this paper proposes the use of data mining techniques to reduce the digital gap. The major data mining techniques considered in this paper are [6]

- a. Association Techniques.
- b. Classification techniques.
- c. Clustering Techniques.

Association: It is method for discovering interesting relations between the variables in the large database. There are different types of algorithm for association rule. They are Apriori algorithm, éclat algorithm, FP-growth algorithm, One-attribute-rule algorithm, Opus search algorithms, and Zero-attribute-rule algorithm [6]. Let us consider the existing E-Governance agriculture database as an example. Suppose, when a user needs a land for the cultivation process with the following features, i.e, good water, larger area, good manpower, and good soil. Based on the above features, the end user can easily search the availability of lands form the existing database with the help of some association algorithm. The one of the best algorithm for technique is Apriori Algorithm.

Classification: It is one of the data mining techniques used to predict the group for data instance. Some of the popular classification techniques are decision trees and neural networks [6]. From the existing database, the end user can classify the land with required parameters like state wise, of district wise, area wise and etc by means of tree like structure. By this classification technique, the user can easily classify the required data from the existing database using some protocols. Based on this, the user can identify the locations and nature of the land with a faster manner. Some of the best and easiest algorithms are decision tree and nearest neighbor algorithm that is available in data mining techniques for better classification.

Clustering: It defined as collection of data object that are similar to one another within the same cluster and dissimilar to the objects in the other cluster. Clustering algorithms are broadly classified into hierarchical and partitioning clustering algorithm (Jain and Dubes, 1988). Again, the Hierarchical algorithm are Agglomerative and Divisive algorithm and the Partitioning Algorithms are k-means, k-mediod, DBSCAN, CLARA, CLARANS, BIRCH CLIQUE, OPTICS etc [6]., When a person is willing to find the group of land for cultivation respective of location, the user can apply the clustering techniques

with the existing e-governance database to form a new groups based upon the user requirement. Thus the user may satisfy.

This is the appropriate time for us to discuss the effectiveness and reachability of technology to all sectors of people. Even the most learned people are reluctant in using the E-Governance technology. This digital divide gap leads to improper usage of Information and Communication technologies. By using the above specified data mining techniques, the digital gap is reduced which in turn help the state to move towards implementing better and quality of E-Governance projects.

3. Improving quality of bandwidth/parameters for better e-governance:

In general, some of the service providers like BSNL, AIRTEL, etc., are available for network connectivity in Tamil Nadu for good quality of Bandwidth. Bandwidth is defined as amount of data transferred in a given period of time [8]. Since, each service providers are having different qualities of bandwidth. But the expected quality of bandwidth is less than the assured bandwidth. As result the network connectivity in Tamil Nadu reached towards down state. Due to this, the successful E-Governance projects get failed while performing data transactions. By considering the above facts, the quality of service (QoS) need to be improved and also all the service provides are expected to provide guarantees for constant network connections. Bandwidth is one of the major constrain for better E-Governance. Some of the parameters are identified to rectify the poor bandwidth problem. For constant connectivity and the better usage of e-governance, the identified parameters are as follows [8]

- a. Availability
- b. Throughput
- c. Data latency
- d. Error rate
- e. Network Traffic
- f. Routing Performance.

Availability: It is defined as the probability that a device will perform a required function without failure under defined conditions for a defined period of time. In most of the case, availability is an important characteristic of system but it becomes more critical and complex issues on networks. With the help of Data Mining technique the network availability are classified with various parameters and helps the service provided for better network availability. Thus, by applying the classification techniques in network database, availability problem will be rectified.

Throughput: It is defined as the rate of communication links or network access. The Throughput is generally measured in bits per second, and sometimes in data packets per second or data packet per time slot. By applying the data mining association algorithm, the service provided will come to normalize the size of the packet for data transformation from one place to another with respect to time and network availability. Based on the mining techniques, the problems are identified and help in future that is not repeated.

Data latency: It is defined as how much time it takes for a packet of data to transfer form one destination point to another destination. The latency mainly depends on the nature of the electromagnetic signal.

Thus the latency may be differing from device to device. Hence, data mining association techniques are applied on the history dataset to identify when the problem happens and how the problem happens; Is it happen previously? If yes, what actions are taken to solve the problem?

Error Rate: It is defined as the number of received bits that have been altered due to noise and interference while during digital data transmission. The error rate may vary from device to device and software application to application. Thus by applying clustering techniques, the service provider can mine the error rate with respect to the hardware and application software from the previous data. Based on this method, the service provider knows which application software Vs hardware device suppose to minimize the error rate.

Network Traffic: It is defined as the data in a network, where the network traffic controller controls the traffic, bandwidth, prioritizing the data packet while during transformation form one point to another. The major part is to measure the network traffic like where the network congestion happens, with this, the classification techniques are applied and the same issue was happened in the previous days or not. Based on the result, the identified problems are rectified.

Routing Performance: It is defined as measuring the performance of the router depends upon the load offered of it, i.e. by means of heavy load of test traffic will reveal the performance. Based on the traffic and load the performance may vary. For better performance, the traffic should be shaped and the packet size should be constant throughout the entire process. With the help of data mining classification techniques, the provider can mine the lesser traffic network for better routing performance.

Network problem happens not only due to technical side but also due to natural calamities, breaking of cable, etc. From the above scenario, the Central or State Government has to rework on the above-mentioned areas to improve the bandwidth performance by means of advanced networking technology, Fiber Optic and recent computing technologies will acted as catalyst for improving bandwidths.

In this paper, bandwidth parameters are analyzed with the help of few data mining techniques for network connectivity to improve the bandwidth. The framework is developed to provide better network connectivity for E-Governance. This paper analyses the bandwidth parameters using data mining techniques and suggest a better framework for improving bandwidth across Tamil Nadu.

4. Proposed frame work for reducing the digital gap:

In Tamil Nadu, there are more successful E-Governance projects being implemented. But, all the implemented applications are heterogeneous in nature *i.e.* the databases are not linked for effective usage. Due to the non-linking of databases and availability in different geographical locations, there exist a digital gap. In the proposed framework, a new concept is introduced to reduce this digital gap, instead of storing the data in different location. This paper proposes the creation of data warehouse, which is a subject-oriented, integrated, time-varying, non-volatile collection of data [5][7].

All the existing and emerging E-Governance databases which is heterogeneous in nature and available in geographical locations are combined and get stored in a common place called 'Data Warehouse'. It may be called as state data warehouse or data repository. The users using a particular E-governance application is able to use the other application also effectively thereby the usage of E-governance applications are increased. Thus, digital gap is also reduced.

From the above figure, the E-Governance technology/applications data are collected from different locations and get stored in different database. This paper proposed a framework in which all the heterogeneous databases are combined and stored in one common place called data warehouse [5]. It contains the summary of all the data, which are made available in a day today process. As per this concept, any one can access any kinds of data at any time by the data warehouse with a faster way. Different data mining techniques are also made available in the proposed framework. By applying these data mining techniques based on the user requirement, the user can mine the data with meaningful order, proper format and in time [7]. Hence, the Tamil Nadu Government E-Governance projects are used more effectively than other State Government projects. With this work, the technology gap is also reduced and the users may utilize the E-Governance by higher level.

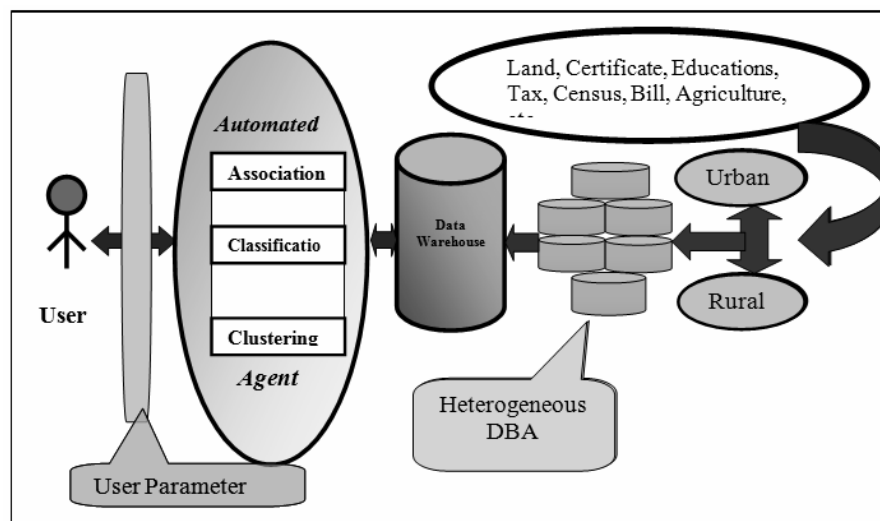


Fig-1: Frame Work for the Betterment of E-Governance.

Conclusion:

In this paper, the importance of digital gaps and the parameters for reducing the digital gap with the E-Governance in Tamilnadu are discussed using different data mining techniques for the better performance. Various network Quality of Services(QOS) parameters such as availability, throughput, data latency, error rate, network traffic and routing performance are considered in data mining perspective to increase the available bandwidth. With the help of proposed framework, the gap also gets reduced between the user and the e-governance systems which enable the Tamilnadu government to implement successful projects.

REFERENCES:

1. Prof. T.P. Rama Rao, "ICT and e-Governance for Rural Development", Governance in Development: Issues, Challenges and Strategies organized by Institute of Rural Management, Anand, Gujarat, December, 2004.
2. Bhatnagar S.C., "E-Government : From Vision to Implementation - A Practical Guide with Case Studies", SAGE Publications Pvt. Ltd., New Delhi, 2004.

3. Rama Rao, T.P., Venkata Rao, V., Bhatnagar S.C., and Satyanarayana J., "E- Governance Assessment Frameworks", <http://egov.mit.gov.in>, E-Governance Division, Department of Information Technology, May 2004.
4. Lee, C.-H., Lee, G.-G., Leu, Y. "Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning", (2009) *Expert Systems with Applications*, 36 (2 PART 1), pp. 1675-1684.
5. Bai, S.-M., Chen, S.-M. "A new method for automatically constructing concept maps based on data mining techniques (2008) " *Proceedings of the 7th International Conference on Machine Learning and Cybernetics, ICMLC*, 6, art. no. 4620937, pp. 3078-3083.
6. Witten I.H., Frank E. "Data Mining: Practical Machine Learning Tools and Techniques". 2nd ed., Elsevier, Morgan Kaufmann Publishers, (2005).
7. Piatetsky- Shapiro, G and Frawley, W.J, " *Knowledge Discovery in Database*," AAAI/ MIT Press, 2000.
8. WWW.compnetworking.about.com, April 2010.

தமிழ் மரபு சார்ந்த தகவல்களின் தகவல் வங்கி, மின் நூல்கள், ஓலைச்சுவடிகளின் ஒருங்கிணைக்கப்பட்ட

இணைய அட்டவணை

சுபாஷினி ட்ரெம்மல்

Technical Consultant , Hewlett Packard, Germany.

Email: ksubashini@gmail.com

துணை தலைவர், தமிழ் மரபு அறக்கட்டளை [http://www.tamilheritage.org]

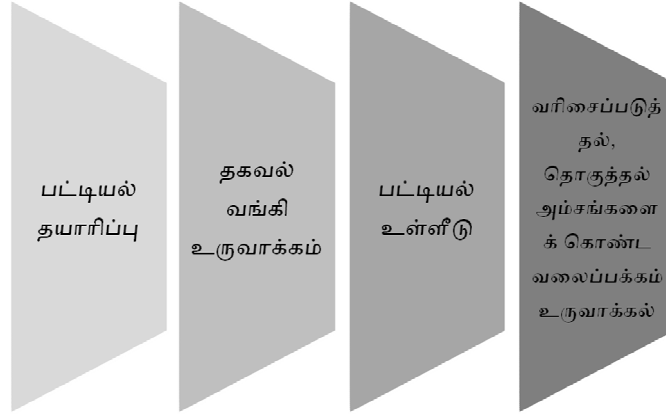
கணிப்பொறியில் தமிழ் பயன்பாடு கடந்த சில ஆண்டுகளில் பன்மடங்கு வளர்ச்சியைப் பெற்றுள்ளது. வலைப்பூக்கள், மடலாடற்குழுக்கள், வலைப்பக்கங்கள் என்பனவற்றோடு ஓர்க்குட், ஃபேஸ்புக் போன்றவை தமிழ் மொழியிலும் இணையத்தில் செய்திப் பரிமாற்றம், செய்திப் பகிர்வு மேற்கொள்ள வகை செய்கின்றன. தமிழ் மொழி சார்ந்த கணினி பயன்பாடு என்பது வெறும் கருத்துப் பரிமாற்றம் என்ற அளவில் மட்டும் நின்று விடாது பல்வேறு தமிழ் மொழி வளர்ச்சி சார்ந்த திட்டங்களை உருவாக்கி இணையத்தில் அதனை பொது பயன்பாட்டிற்குத் தரும் வகையிலும் பல முயற்சிகளைக் கண்டுள்ளது. இணைய உலகில் கணினி பயன்பாடு வழங்கியிருக்கும் மகத்தான வாய்ப்பினைப் பயன்படுத்தி தமிழ் மொழி வளர்ச்சி, பாதுகாப்பு சார்ந்த பல்வேறு நடவடிக்கைகளை மேற்கொள்ள வாய்ப்புக்கள் பெருகி வருகின்றன. தமிழ் மொழி வளர்ச்சி என்பது புதிய இலக்கியங்களின் உருவாக்கங்கள் மட்டுமின்றி பழம் தமிழ் நூல்கள், கல்வெட்டுக்கள், ஆவணங்கள், ஓலைச்சுவடிகள், வரலாற்று ஆவணங்கள் போன்றவற்றை மின்வெளியில் பாதுகாப்பதிலும் அடங்கும். தமிழ் மொழியிலுள்ள தகவல் வளங்களைத் திரட்டி இணையத்தில் வெளியிடுவது இவ்வகைத் தகவல் தேடுபவர்களுக்கு மட்டுமன்றி இணையத் தொடர்பு உள்ள அனைவரும் வாசித்துப் பலனடையவும் வாய்ப்பளிக்கும். இதனைக் கருத்தில் கொண்டு தமிழ் மரபு அறக்கட்டளை இணையத்தில் பழம் தமிழ் மின்னூல்கள் பட்டியல், ஓலைச் சுவடி அட்டவணை, கல்வெட்டு பட்டியல் அட்டவணை ஆகியவற்றை உருவாக்கும் முயற்சிகளில் ஈடுபட்டு வருகின்றது.

தமிழ் மரபு அறக்கட்டளை எனும் தன்னார்வத் தொண்டுமிய நிறுவனம் 2001ம் வருடம் அதிகாரப்பூர்வமாக உருவாக்கப்பட்ட உலகு தழுவிய ஒரு இயக்கமாகும். கணினி தொழில் நுட்பத்தைக் பயன்படுத்தி ஓலைச்சுவடிகளை மின்பதிவாக்கி அதனை வாசிப்பிற்கும் ஆய்விற்கும் உட்படுத்த முடியும் என்ற வகையில் சிந்தனையை வளர்த்து வரும் ஒரு பேரியக்கம் இது. இந்த நிறுவனம் ஓலைச் சுவடிகள் மட்டுமல்லாது மறுபதிப்பு காணாத நூல்களும் கூட அழியக் கூடிய சாத்தியம் உள்ளது என்பதைக் கருத்தில் கொண்டு மின்னூலாக்க முயற்சிகளில் ஈடுபட்டு வருகின்றது. தற்போது தமிழ் மரபு அறக்கட்டளை தமிழ் மரபு சார்ந்த தகவல்கள் சேகரிப்பு, இணையத்தில் இவ்வகைத் தகவல்களைப் பதிப்பித்தல் அத்துடன் அவற்றை பொதுமக்கள் வாசிப்பிற்கு இணைய தொழில் நுட்பத்தை பயன்படுத்தி தகுந்த முறையில் வெளியிடுதல் என்ற பணிகளில் தொடர்ந்து ஈடுபட்டு வருகின்றது. இந்த நிறுவனத்தின் அதிகாரப்பூர்வ வலைப்பக்கத்தை <http://www.tamilheritage.org/> என்ற பக்கத்திலும் இந்த அமைப்பின் மின்தமிழ் மடலாடற்குழுவை <http://groups.google.com/group/minTamil> பகுதியிலும் காணலாம்.

மின்பதிப்பாக்கம், மின்னூலாக்கம் ஆகியற்றின் தொடர்ச்சியாக இணையத்தில் மின் நூல்களுக்கான அட்டவணை ஒன்று உருவாக்க வேண்டிய தேவை உள்ளது என்பதைக் கருத்தில் கொண்டு முதல் சோதனை

முயற்சியாக தமிழ் மரபு அறக்கட்டளை பழம் மின்னூல்கள் பட்டியல் ஒன்றினை உருவாக்கும் முயற்சியைத் தொடங்கியது.

முதலில் வெளியிடப்பட்ட பட்டியல் html அடிப்படையில் உருவாக்கப்பட்ட ஒரு பக்கம். இந்த பக்கம் ஒருங்குறி தமிழ் எழுத்துருவில் மின்னூல்கள் பட்டியல் உள்ளீடு செய்யப்பட்டு உருவாக்கப்பட்டுள்ளது. இந்த பக்கத்தில் உள்ள நூல்களைத் தேடு இயந்திரத்தின் மூலம் தேடும் வகையில் எழுத்துக்கள் படக்கோப்புக்களாக இல்லாமல் ஒருங்குறியில் அமைக்கப்பட்டுள்ளன. இந்த அடிப்படை வலைப்பக்கத்தில் நூல்களின் பெயர்களையோ, நூலாசிரியர் பெயரையோ அல்லது பதிப்பித்தோர் பற்றிய தகவலையோ வரிசைப்படுத்தி பார்க்க முடியாது. வழங்கப்பட்ட தகவலை பட்டியலில் உள்ளவாறு காட்டும் வண்ணம் இந்தப் பக்கம் அமைக்கப்பட்டுள்ளது. அடிப்படையாக அமைந்த இந்தப் பக்கத்தை மேம்படுத்தி மேலும் வரிசைப்படுத்துதல், ஒரே ஆசிரியரின் நூலைத் தேர்ந்தெடுத்தல், ஒரே வகையான நூல்களைப் பட்டியலிடுதல், ஒரே ஆண்டில் வெளிவந்த நூல்களைப் பார்வையிடுதல், ஒரு நூல் ஆசிரியரின் பிற நூல்களைப் பட்டியலிடுதல் போன்ற சில சிறப்பு அம்சங்களை இந்தப் பக்கத்தில் சேர்ப்பது இப்பக்கத்தைப் பார்வையிடுவோருக்கும் பயன்படுத்துவோருக்கும் உதவும் என்ற கருத்தில் இவ்வகை சிறப்பு அம்சங்களை நுழைக்க முயற்சித்தோம். இதற்கு நான்கு படி நிலைகள் அடையாளம் காணப்பட்டன:



இந்த 4 படி நிலைகளின் அடிப்படையில் உருவாக்கப்படும் வலைப்பக்கத்தை உருவாக்க கீழ்க்காணும் தயாரிப்பு நடவடிக்கைகள் மேற்கொள்ளப்பட்டன.

1. இணையத்தில் ஏற்றப்பட்ட நூல்களின் இணைப்புக்கள் அனைத்தும் சரியாக இயங்குவது உறுதி செய்தல்
2. பட்டியலைத் தயாரித்தல்
3. பட்டியலுக்காக ஒரு சிறப்பு தகவல் வங்கியை செர்வரில் உருவாக்கல்
4. தகவல் வங்கியில் தொடர்பு கொள்ளவும் அதனைச் செயல்படுத்தவும் admin பயனாளரை உருவாக்குதல்
5. தகவல் வங்கியில் பட்டியலில் உள்ள தகவல்களைச் சேர்க்க ஒரு குறிப்பிட்ட பகுதி (table) ஒன்றினை உருவாக்கி அதற்குத் தேவையான அனைத்து தவல்களையும் வழங்கி தேவையான table structure உருவாக்கல்
6. உருவாக்கப்பட்ட பட்டியலைத் தகவல் வங்கியில் உருவாக்கப்பட்டுள்ள பகுதியில் (table) அதனை இணைத்தல் (data import)
7. php அடிப்படையில் தயாரிக்கப்பட்ட வலைப்பக்கத்தை உருவாக்கி பட்டியலை வெளியிடுதல்.

இந்தத் தயாரிப்பு நடவடிக்கைகளின் அடிப்படையில் சோதனை முயற்சி தொடங்கப்பட்டது. இதுவரை தமிழ் மரபு அறக்கட்டளையின் பல்வேறு திட்டங்களின் வாயிலாக மின்னாக்கம் செய்யப்பட்ட மின்னூல்களின் வலைப்பக்க இணைப்புக்கள் சோதிக்கப்பட்டன. உடைந்த இணைப்புக்கள் சரிபார்க்கப்பட்டு மீண்டும் சரியான வலைப்பக்க முகவரி ஒவ்வொரு நூலுக்கும் வழங்கப்பட்டது. இதனைத் தொடர்ந்து ஒவ்வொரு நூலுக்குமான தகவல்கள் ஒரு excel spreadsheet ஒன்றில் உள்ளீடு செய்யப்பட்டன. உதாரணமாக மின்னூல் எண், நூலின் பெயர், நூலின் ஆசிரியர், நூல் வெளிவந்த ஆண்டு, பதிப்பாசிரியர், மின்னூல் உருவாக்கியவர் பெயர், மின்னூலின் தன்மை போன்ற தகவல்கள் இந்தப் பட்டியலில் ஒவ்வொன்றாக இணைக்கப்பட்டன. இந்த excel spreadsheet கோப்பு *.csv வகையில் சேகரிக்கப்பட்டது.

இதற்கடுத்தார்போல் தமிழ் மரபு அறக்கட்டளை செர்வரில் MySQL தகவல் வங்கி மென்பொருளைக் கொண்டு ஒரு பிரத்தியேக தகவல் வங்கி உருவாக்கப்பட்டது. (குறிப்பு: தமிழ் மரபு அறக்கட்டளை செர்வர் Linux இயங்கு தளத்துடன் Apache, PHP மற்றும் MySQL 5.0 சேர்த்து உருவாக்கப்பட்ட ஒரு web server.) அதில் தகவல்களைச் சேர்த்து வைக்க ஒரு பிரத்தியேக பகுதி (table) ஒன்றும் அதற்கான வடிவமைப்பும் (structure) அமைக்கப்பட்டன. அதிலும் குறிப்பாக முதலில் தயாரித்த excel spreadsheet கோப்பில் உள்ளவாறு தகவல் தொகுப்புக்கள் (fields) உருவாக்கப்பட்டன. இத்தகவல் வங்கி மேற்பார்வைக்காக ஒரு பிரத்தியேக பயணாளர் கணக்கும் உருவாக்கப்பட்டது. இதற்கடுத்து உருவாக்கப்பட்ட தகவல் வங்கியில் உள்ளீடு செய்யப்பட்ட தகவல்களை ஏற்ற வேண்டும். இதற்கு செர்வரில் MySQL தகவல் வங்கிக்காக பிரத்தியேகமாக cpanel-லில் இணைக்கப்பட்டுள்ள PhpAdmin மென்பொருள் கொண்டு ஏற்கனவே உள்ளீடு செய்யப்பட்ட excel spreadsheet கோப்பினை கணினியிலிருந்து செர்வருக்கு ஏற்றப்பட்டன. இந்த முறையில் இந்த பிரத்தியேக தகவல் வங்கியில் மின்னூல்கள் அனைத்திற்குமான தகவல்கள் ஏற்றப்பட்டன. அடுத்த கட்டமாக ஒரு வலைப்பக்கத்தின் வாயிலாக இப்பட்டியலைப் படிப்பதற்கு ஏதுவாக தனிப்பக்கம் ஒன்றினை உருவாக்க வேண்டும்.

இதனைச் செய்வதற்கு அடிப்படையில் php அடிப்படையாகக் கொண்ட வலைப்பக்கம் உருவாக்கப்பட்டது. இந்தப் பக்கம் நேரடியாக செர்வரில் உள்ள இந்தப் பிரத்தியேக தகவல் வங்கியை தொடர்பு கொண்டு, தொடர்பு தகவல்கள் சரியாக உள்ளனவா என்று உறுதி செய்யப்பட்டவுடன் அதில் இணைக்கப்பட்டுள்ள தகவல்களை இந்தப் பகுதி வாசிக்க வேண்டும். வாசித்து இந்த பக்கத்தில் கொடுக்கப்பட்டுள்ள கட்டளைகளின் அடிப்படையில் தகவல்களை வரிசைப்படுத்தி ஒரு தனி வலைப்பக்கத்தில் தகவல்களைப் பட்டியலாக்கிக் காட்ட வேண்டும். இந்த பக்கத்தில் பட்டியலில் உள்ள நூல்கள் அகர வரிசைப்படுத்தவும், ஒரே தலைப்பிலான நூல்களைத் தேட, மற்றும் ஒரே ஆசிரியரின் நூல்களைத் தேட, ஒரே பதிப்பாசிரியரின் நூல்களைத் தேட என பட்டியல் தேடல்களுக்கும் வரிசைப்படுத்தலுக்கும் உகந்தவாறு உருவாக்கப்பட்டது.

முதல் சோதனை முயற்சியின் அடிப்படையில் கல்வெட்டுக்களின் பட்டியல் அடங்கிய தொகுப்பு ஒன்றினை மின்பதிப்பக்கம் செய்ய முனைந்தோம். இதற்கு ஆதாரமாக தொல்பொருள் ஆய்வுத்துரை அறிஞர் முனைவர்.ஆர்.நாகசாமி அவர்களது "உங்கள் ஊர் கல்வெட்டுத் துணைவன்" Pathway to the Antiquity of your Village" எனும் ஆய்வு நூலை இந்தப் பணிக்கு பயன்படுத்தினோம். (நன்றி: திருமதி.கீதா சாம்பசிவம்,சென்னை. - நூலில் உள்ள அனைத்து தகவல்களையும் முழுதும் தட்டச்சு செய்து அனுப்பி வைத்தவர்). இந்தப் பட்டியலைத் தயாரிக்க மின்னூல் பட்டியலுக்கு செய்த ஏற்பாடுகளைப் போலவே ஒரு பிரத்தியேக தகவல் வங்கி உருவாக்கப்பட்டு இந்த ஆன்லைன் அட்டவணை அடங்கிய பக்கம் உருவாக்கப்பட்டது.

சுவடிகளுக்கான பட்டியல் இதுவரை இணையத்தில் பதிப்பிக்கப்படவில்லை. அச்சு வடிவத்தில் உள்ள பனை ஓலைச்சுவடிகளின் பட்டியல் இணையத்தில் பதிப்பிக்கப்பட வேண்டியதன் அவசியம் உள்ளது. இதுவரை சுவடிகளுக்கான பேரட்டவணை சில அச்சுவடிவத்தில் வெளிவந்துள்ளன. தொடர்ந்து பல நிறுவனங்கள் சுவடிகளின் அட்டவணையைச் சேகரிக்கும் முயற்சிகளில் இயங்கி வந்தாலும் இதுவரை

சேகரித்துள்ள அனைத்து சுவடிகளுக்குமான முழுமையான ஒரு பட்டியல் இணையத்தில் இல்லாதது கவனத்தில் கொள்ளப்பட வேண்டிய ஒன்று. இதனைக் கருத்தில் கொண்டு இந்த நடவடிக்கை தொடங்கப்பட்டுள்ளது.

இவ்வகை அட்டவணைகள் ஆய்வாளர்களுக்கு மட்டுமல்லாமல் பொதுவான வாசிப்பிற்கும், தகவல் தேடுபவர்களுக்கும் மிக எளிதாக தகவலை வாசிக்க மிக உதவும். இதனைக் கருத்தில் கொண்டு தொடர்ந்து இணையத்தில் இதுவரை சேகரிக்கப்பட்ட பணை ஓலைச்சுவடிகளின் அட்டவணையை இணையத்தில் உருவாக்க இப்போது முயற்சியில் தமிழ் மரபு அறக்கட்டளை ஈடுபட்டுள்ளது. இதற்கு ஆதாரமாக தஞ்சாவூர் தமிழ்ப்பல்கலைக்கழகத்தின் வெளியீடுகளான தமிழ் ஓலைச்சுவடிகளின் பட்டியல் இந்த திட்டத்திற்கு பயன்படுத்தப்பட உள்ளது.

An Extended Cross Lingual Information Retrieval System for Agricultural Domain using Statistical Document Translation for Tamil Farmers

D. Thenmozhi, Arun Balachandran Ganesan and C. Aravindan

Department of Computer Science & Engineering

SSN College of Engineering, Chennai, India

theni_d, arunbalachandrang, aravindanc@ssn.edu.in

Abstract

Cross Lingual Information Retrieval (CLIR) system allows a user to pose a query in one language and search documents in a different language. In this paper, we extend the use of Tamil-English CLIR system for Tamil farmers by translating the English web pages to Tamil, through which the farmer can pose a query in Tamil and read the information in the same language. We have developed statistical English-Tamil translation engine to translate the English documents that are retrieved from CLIR system to Tamil. The parallel corpus for this statistical engine was built using the text in the domain of Agriculture for Tamil and English languages. The translation model is then trained with this sentence aligned domain-specific corpus. Also a text corpus for Tamil has been built and used in building and training the language model. A Statistical Machine Translation Decoder tool has been used to perform the decoding as the final step.

Introduction

The world of information is huge and expanding; also most of the information is available in English. The non English speaking users still find it a major problem in utilizing this vast resource of information. Accessing this information through queries written in Indian languages is still more difficult. CLIR systems solves this problem by allowing the non English user to specify their information need in their native language and accessing rich information that are available in English.

We have developed a Tamil-English CLIR system for Agriculture society [8] which allows the Tamil farmers to pose Tamil queries and retrieves information from English corpus. The CLIR systems generally display the search result in English. It is appropriate, if the results are displayed in their own language for the users who do not know how to give query in English.

In this paper, we extend the CLIR system by translating the retrieved English documents to Tamil in the domain of Agriculture through which Tamil farmers can enter their query in Tamil, retrieves relevant pages from English corpus and read the information in Tamil.

Machine translation is a sub-field of computational linguistics that investigates the use of computer software to translate text from one language to another. It can use a method based on linguistic rules, when the content to be translated is simple, for example short queries. When the content to be translated is complex in structure, machine translation requires the problem of natural language understanding to

be solved first. Statistical machine translation solves this problem with a probabilistic approach based on bilingual text corpora.

We have built English-Tamil parallel corpus in the domain of Agriculture to train the statistical translation system for translating the retrieved documents in English to Tamil.

Our translation system may contribute in developing Tamil Wikipedia by translating the Wikipedia content that are available in English in the domain of Agriculture to Tamil.

Literature Survey

a. Cross Lingual Information Retrieval System

The Advanced Cross-Lingual Information Access (ACLIA) currently works on two tasks Information Retrieval and Question Answering [7]. For both these tasks, they work for cross-lingual and mono-lingual topics for the languages English (EN), Simplified Chinese (CS), Traditional Chinese (CT), and Japanese (JA) for the tracks EN-CS, EN-CT, EN-JA (cross-lingual) and CS-CS, CT-CT, JA-JA(mono-lingual). [7] created the Evaluation Package for ACLIA and NTCIR (EPAN). The EPAN toolkit contains a web interface, a set of utilities and a backend database for persistent storage of evaluation topics, gold standard nuggets, submitted runs, and evaluation results for training and formal run datasets. [4] developed a Question Answering system that answers complex questions from multilingual sources. They improved the performance of Chinese-to-Chinese and Japanese-to-Japanese subtasks of NTCIR7 IR4QA by means of Mean Average Precision. ACLIA presently not working on the document translation of cross-lingual track.

Many research groups are working on the CLIR system for Indian Languages [6]. [5] developed a Tamil-English CLIR system that uses Tamil morphological analyzer for language analysis. A named entity recognizer is used to identify the named entities in the document for indexing and ranking. They used a bi-lingual dictionary approach for translation. A statistical engine based on n-gram approach is used for transliteration. A simple ontology is used for query expansion. Ranking is based on term frequency. They evaluated the system for the news domain by collecting the English corpus from the magazine "The Telegraph".

We differ from the above works by extending Tamil-English CLIR system with document translation in the domain of Agriculture which helps Tamil farmers who have English as a language barrier.

b. Machine Translation System

[1] developed a English to Tamil translation system using statistical approach. SRILM toolkit it used for language modeling. Translation model is trained using the parallel corpus in the News domain. They learnt Named entities from this statistical machine translation system. Phrase based decoder is used to translate the English sentences to Tamil.

[3] used the same phrase based approach of statistical machine translation to transliterate word from English to Hindi, Tamil and Kannada languages for the Named-entities. They used GIZA++ for word alignment and Moses for decoding.

[2] improved the English-Hindi statistical machine translation system by considering the case markers and morphology of Hindi language. They used factored based translation model instead of phrase based model.

We also developed a similar kind of statistical machine translation system that translates English sentences to Tamil by using SRILM toolkit for the language modeling of Tamil and Moses for translation modeling and decoding.

System Architecture

The extended CLIR system uses a three phase model that accepts the query from the user in Tamil extracts the documents from English corpus and translates them to Tamil in the domain of Agriculture. This is illustrated in the figure 3.1.

a. Tamil-to-English Query Translation

A rule based Tamil-to-English translation engine is developed through which the query is entered in Tamil and translates to English. It uses morphological analyzer to split the query into individual words. These words are translated to English using a Tamil-English bilingual dictionary. When Tamil words gets ambiguous meaning, the correct meaning is obtained based on the context using word sense disambiguator. The obtained English words are re-ordered based on the syntactic structure of English language (Subject-Verb-Object pattern).

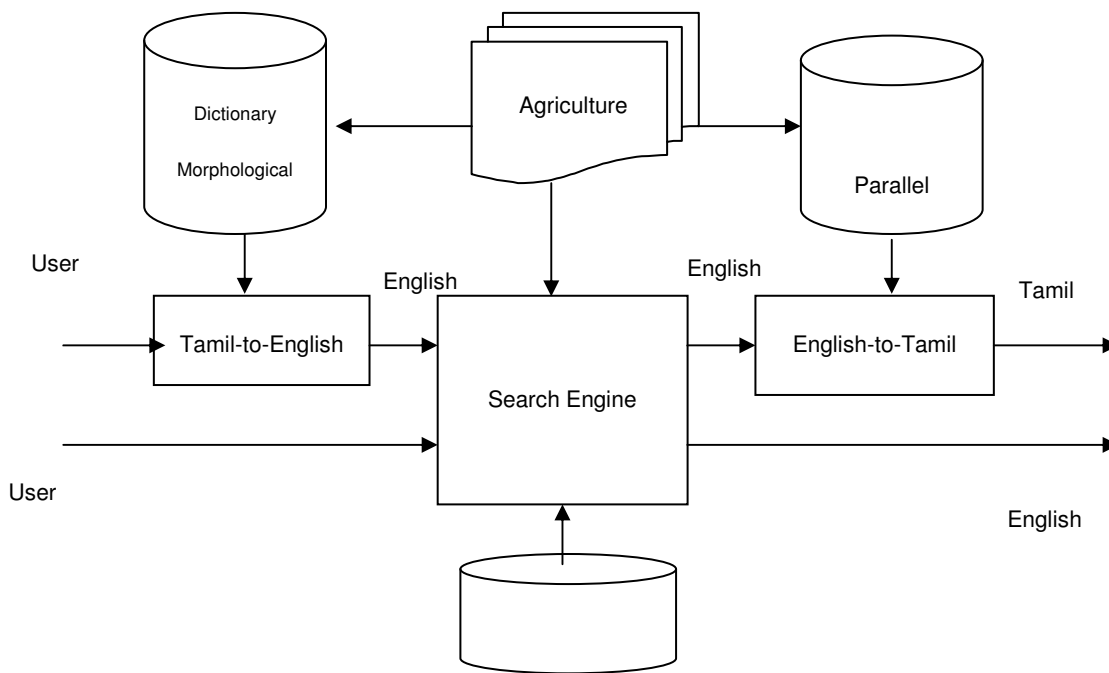


Figure 3.1. System Architecture

b. Searching and Retrieval

The translated English query is given to the existing search engines like Google, AltaVista, etc that retrieves the English documents. We considered top twenty pages that are retrieved by the search engine for statistical document translation. The documents with the extension doc, html and pdf are converted to plain text before given to the English-to-Tamil translation Engine.

c. English-Tamil Document Translation

A statistical English-Tamil document translation engine is developed with the phases namely constructing English-Tamil parallel corpus, training the translation model with the parallel corpus, training the language model for Tamil language and decoding.

Tamil Wikipedia, and the Tamil Nadu Agricultural University websites have been identified as excellent sources that would help in providing the data needed for corpus building. Information in documents has been extracted/retrieved using a semi-automated text extractor and unnecessary attributes to the data is removed either by automation or human intervention depending on the type of attribute. The encoding of the content in Tamil language is converted to Unicode standard. The aligned parallel corpora that is built follows Unicode standard.

Moses toolkit is used to train the translation model using the parallel corpus based on phrase based approach. This model works on the principle of Baye's rule

$$P(t|e) = P(e|t) * P(t) / P(e)$$

$$T = \text{argmax } P(e|t) * P(t)$$

where T is the translation probability.

SRILM toolkit is used to build tri-gram language model for Tamil language. Tamil sentences extracted for building parallel corpus in the domain of Agriculture is used to build this model. A beam search decoder of Moses toolkit for phrase-based statistical machine translation model is used to translate the sentences of English documents to Tamil.

Experimental Evaluation

a. Agriculture Corpus

We extracted the parallel sentences from Tamil Nadu Agriculture University (TNAU) website and Wikipedia content in the domain of Agriculture. Some additional sentences in Tamil from the TNAU website is used to train the language model. The number of sentences used to train the translation model and language model are given the table.

	Number of Sentences	
	TNAU Site	Wikipedia
Translation model	6135	226
Language model	6900	-

b. Comparison with Google Translation System

We compared the performance of our extended system with the Google translation system by various measures that are listed in the following table.

	Google Translation System	CLIR system
Tamil Queries	Transliterated Approach	Machine Translation Approach
Document Retrieval for Tamil Queries	From Tamil Corpus	From English Corpus
Recall value for Tamil Queries	Less	More
Translation	Available for many Asian/ European languages. Tamil is not available	Translates English documents to Tamil

Conclusion

This extended CLIR system helps the Tamil farmers to look for rich information available in English corpus by specifying their need in Tamil and view the results also in Tamil language. This system uses a rule based approach for query translation and statistical based approach for document translation. With the available parallel corpus, query translation also can be done statistically in future. Factored based translation model can be used for improving document translation. Query also can be enriched with more semantics by using Agriculture ontology.

References

1. Amrita Vishwa Vidyapeetham, "valluvan -English to Tamil Statistical Machine Translation", Center for Excellence in Computational Engineering and Networking (CEN), 2005
2. Ananthakrishnan Ramanathan, Hansraj Choudhary, Avishek Ghosh, Pushpak Bhattacharyya, Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT, Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp 800-808
3. Manoj Kumar Chinnakotla and Om P. Damani, Experiences with English-Hindi, English-Tamil and English-Kannada Transliteration Tasks at NEWS 2009, Proceedings of the 2009 Named Entities Workshop, ACL-IJCNLP 2009, pp. 44-47
4. Ni Lao, Hideki Shima, Teruko Mitamura, and Eric Nyberg, "Query Expansion and Machine Translation for Robust Cross-Lingual Information Retrieval", Proceedings of NTCIR-7 Workshop, 2008, Japan
5. Pattabhi R.K Rao and Sobha. L, "AU-KBC FIRE2008 Submission - Cross Lingual Information Retrieval Track: Tamil- English", First Workshop of the Forum for Information Retrieval Evaluation (FIRE), pp 1-5, 2008
6. Prasenjit Majumder, Mandar Mitra Swapan parui and Pushpak Bhattacharyya, "Initiative for Indian Language IR Evaluation", Invited paper in EVIA 2007 Online Proceedings.
7. Teruko Mitamura, Eric Nyberg, Hideki Shima, Tsuneaki Kato, "Overview of the NTCIR-7 ACLIA Tasks: Advanced Cross-Lingual Information Access", Proceedings of NTCIR-7 Workshop, 2008, Japan
8. Thenmozhi D, Aravindan. C, "Tamil-English Cross Lingual Information Retrieval System for Agriculture Society", 8th International Conference on Tamil Internet, Germany, 2009

Tamil E-Archives Management

B. Asadullah

Librarian

Mazharul Uloom College, Ambur – 635 802.

Vellore District, Tamil Nadu.

Abstract

From the year 2000 onwards, we had noticed that new innovations were made enormously to serve the human needs in every core area from Agriculture to Engineering, Medical, and Automobile etc., from every corner of the universe in every language of the world, and more especially by Tamil Literature. This has compelled us to tap all these new findings and old traditional findings in a secure way so that these resources can be better utilized in future by our new generation peoples.

In keeping these concerns in mind, I am trying to explain how we can secure all these findings in a secure way using the suitable technology by applying possible low cost of implementation and maintenance. This paper will explain the way, on how to manage the archives of new findings and old traditional findings of Tamil languages in electronic format using pictorial images and formal text format utilizing the best technology at low cost using best available free of cost Open Source Software.

Introduction

Tamil language is spoken and widely used by over 66 million peoples across the world in particular in three countries as official language and in six countries as monitories, where our state of India Tamil Nadu is the largest user of Tamil in every aspect of life. Tamil scripts are preserved in different format from palm leaf to paper, due to accessibility and timely availability. But this preservation does not last for number of years as every functional item has a stop point that is everything has a counted life span. This article will help in preserving archive in computer using latest and freely available licensed open source software.

Electronic Preservation

Due to moderate and vast development of Computer Technology, preservation has become reliable, secure and durable. Here preservation means not only documents, but for vast range of items in electronic format like running text, document, book, journal, pictorial, audio, video and animated format etc., all these above said format are also known as E-Achieves.

Preservation in computer is also a big question, but the answer to this question is having backup of all thing in a compact format, or having a mirror storage which can be made possible using the RAID technology while storing the data in computers. In case of one's end of life, another mirror image can be utilized in less effort, instead of building whole thing from scratch.

From the above last two clause, we come know that preservation in computer is possible, durable and securable. But maintenance of unlimited e-archives in a computer is not so simple and easily accessible for all end-users.

For this purpose, after wide consultation and analysis, it was concluded that there were a suitable software which fit for the chosen topic, and it is available free of cost in different users operating environment also known as operating system like Microsoft Windows, Linux, under the Open Source Software GNU license.

Institutional Repository

An Institutional Repository is an online locus for collecting, preserving, and disseminating in digital form. (Also known as Digital Library). The main objectives for an institutional repository are to create global visibility for an institution's scholarly research and to collect content in a single location, to provide open access to institutional research output by self-archiving it and to store and preserve other institutional digital assets, including unpublished works. Some examples are Fedora, E-Print, IR+, and DSpace, Green Stone.

DSpace is developed by HP Lab and MIT Libraries of US. At the time of writing this article for presentation at the 9th Tamil Internet Conference held at Coimbatore, Tamil Nadu (India) 2010, DSpace 1.6.2 is available for utilization. DSpace is widely accepted and used by different educational institutions across the world.

DSpace is open source software based on BSD License, it is an digital object / assets management system to create, store, search and retrieve digital objects. It allows open access to digital objects and building institutional repositories and the collections are searchable and retrievable on the web. Since it is open source technology platform which can be customized and its capabilities can be extended.

Main features of DSpace are it is a low cost, including all hardware and software components. Technically simple to install and manage, robust, scalable, open and inter-operable, it is modular and user friendly with multi user environment (including both searching and maintenance), it is multimedia digital object enabled and platform independent interoperable. DSpace stores all data as a digital object, which has the following standard model.

Objects in the DSpace Data Model	
Object	Example
Community	Like Department / Section
Collection	Report, Statistical Data
Item	Technical report, A data set with accompanying description
Bundle	A group of HTML, image bitstream
Bitstream	A single HTML file, a image file or a text file
Bitstream format	MS Word Document, JPEG image format

DSpace has its own Ingest Process and Workflow Steps in three stage, which verifies all type of documents with their corresponding information, then only it accept the document to be stored in the repository. This method helps in gathering authenticated Tamil document for building a healthy achieve.

DSpace uses the CNRI Handle System for creating identifier, which is required by DSpace to store and locate the independent mechanism for creating and maintaining digital objects.

DSpace uses the Lucene Search algorithm, which helps in searching the tamil text using Fielded search, Boolean, Exact term, proximity and wild cards, fuzzy and range, boosting terms search techniques.

Content Management System

A content management system is a collection of procedures used to manage work flow in a collaborative environment. These procedures can be manual or computer based. The procedures are designed to allow for a large number of people to contribute to and share stored data with controlled access to data, based on user roles. User roles define what information each user can view or edit and aid in easy storage and retrieval of data by reducing repetitive duplicate input with improving the ease of report writing and improved communication between users.

In a content management system, data can be defined as nearly anything documents, movies, pictures, phone numbers, scientific data, etc. These are frequently used for storing, controlling, revising, semantically enriching, and publishing documentation. Examples are web content management, digital asset management, digital records management and electronic content management. Synchronization of intermediate steps, and collation into a final product are common goals of each. The features of content management systems are, it can structure the document repository as per requirements and automatic document versioning control with content categorization and role-based access control. Using varied content types, it can be managed like HTML, Unicode Tamil text, PDF, Documents of different standard of any language, image files, video clips etc., Document can be accessed through web-browser and network drive mapping with full text searching facility in txt, doc, pdf, odt, sxw, html, xls, sxc, ods etc.,

Document Repository System

A document repository system is a systematic way of storing documents along with a set of user defined attributes that need to be filled in at the time of storage. Documents can be searched based on these attributes or on their content. Every organization needs a good planned Document repository system in their specific official language like in our case Tamil language.

Owl is one of the best examples for document repository system, Owl is a powerful system, full of use full features, it allows defining personal sets of attributes for documents. MS-Word of Tamil language, PDF, Unicode TEXT of Tamil, can be word indexed and searched. However documents and files of any type can be stored. Owl has security model for document access.

The documents are uploaded and stored in a folder, uploaded files can be stored in tables in the open source database like RDBMS, MySQL, PostgreSQL for storing file / folder related information.

Conclusion

Managing achieve is not the easy task, when large numbers are being created day to day because of vast development of Tamil language and literature. In this context managing archive has become easy with the help of computers using proper software depending on our usage and requirement.

From the above article, we come to conclusion that DSpace is good digital library software, which helps in managing the vast digital objects of Tamil literature by giving good retrieval facility with modern technology.

Content management system, also helps in managing E-Archive, but its method is different when compare with Institutional repositories software. Document repository system, help in storing e-achieves systematically.

References.

- Wilson Katie : Computers in Libraries, New Delhi, New Age International Publications., 2006.
- Bhatt, R K : March towards digitations of Information Resource in India.
- Sreekumar MG and others Ed., Digital Libraries in Knowledge Management, New Delhi, ESS ESS Publishers, 2006.
- <http://www.unesco.org/>
- <http://ta.wikipedia.org/wiki/>
- <http://www.eprints.org/>
- <http://www.irplus.org/>
- <http://www.dspace.org/>
- <http://www.greenstone.org/>
- <http://www.infitt.org/ti2010/register/papersubmission.php>

Very long-term digital preservation and archival strategies for Tamil documents

Mani M. Manivannan

Senior Director of Engineering

Symantec Corporation

Chennai, India

mmanivannan@gmail.com

Abstract

The decipherment of Indus script remains controversial. For several years, the inscriptions in Tamil Brahmi script weren't recognized as Tamil. The inscriptions in Vaṭṭeḷuttu, Pallava grantha, and evolving Tamil Brahmi scripts required careful analysis to decipher. Probable errors in inscriptions and reading of the inscriptions have led to multiple interpretations. As writing instruments and technology changed, Tamils have lost valuable literature and public documents. In the short duration that digital Tamil has existed, we are already finding it difficult to retrieve early Tamil writings on the internet. As we embark on the e-Governance and digital archival of precious documents, we have to remember that unless we are careful, we may find it difficult to preserve these documents for the long duration of centuries. In this paper, we will explore some of the strategies to minimize the loss of valuable documents.

1. Background and Overview

Tamils have been creating digital documents in Tamil since the mid 1980s[1] using MS-DOS based editors. Since the early 1990s Tamils have been privately exchanging e-mails in Tamil and with the advent of tamil.net mailing list in 1996, Tamils worldwide started to communicate with each other in Tamil script. Since then there have been multiple Tamil mailing lists, web pages, webzines, blogs that added Tamil digital content. Project Madurai, a community project has been digitizing classical Tamil literature and Tamil books in the public domain and creating a major Tamil corpus with its digital documents since January 1998. Though the Tamils of the diaspora were instrumental in creating a lot of the early Tamil digital content on the web, the mainland Tamil Nadu media have taken to the web and nearly all the Tamil news media have Tamil content on the web. Tamil Nadu government departments have been creating Tamil digital documents in various encodings including TAB, TAM and Baamini. With a nationwide push for e-Governance in India, the Tamil Nadu government is about to embark on one of the largest Tamil digitization efforts.

Before the advent of the standardization efforts to unify the multiple encodings after the TamilNet'97 conference in Singapore, there were multiple fonts each with its proprietary encoding. These discussions led to the TSCII encoding, the first open, non-proprietary Tamil encoding specification in 1998. At the TamilNet'99 conference in Chennai, the Government of Tamil Nadu formally declared two encoding standards TAB and TAM. During this period, neither ISCII, the Indian national standard and Unicode the global standard based on ISCII, had much support among the Tamil developers. However the standards TSCII, TAB and TAM did not stop others from creating fonts based on proprietary encodings.

Baamini has been popular among the Eezham Tamil diaspora before TSCII and it is still used by some in the Government of Tamil Nadu. Other encodings that were used to create Tamil digital documents include Vanavil, Indoweb, Murasoli, Webulagam, Thinathanthi, Dinamani, Thinaboomi, Murasu Anjal, Mylai, Thatstamil, ShreeLipi, Amudham (Dinakaran), Vikatan, Anu Graphics (Pallavar), and Senthamizh (Nakkeeran). These encodings are popular enough that some converters recognize all of these. With strong support for the Unicode encoding in Microsoft Windows and Apple platforms as well as Google’s search and e-mail applications, use of Unicode has started to spread. Tamil Wikipedia is a popular site that uses Unicode as does the Tamil Lexicon at Chicago University. With the National e-Governance plan of Government of India notifying Unicode as the standard for e-Governance applications, creation of Tamil digital documents in Unicode encoding is expected to increase substantially.

INFITT has recognized Unicode as the main standard for Tamil computing. INFITT has also acknowledged that there are commercial applications that don’t yet support Unicode, such as those used in publishing and other industries. For these applications that are not Unicode-ready, INFITT has recognized Tamil All Character Encoding TACE-16 as the only Alternate Standard for Tamil Computing. Since high-end publications are expected to create PDF documents such as Government notifications, text books, etc., along with books published by vendors in Tamil using TACE-16 encoding, this will form part of the Tamil digital document collection.

In this paper, we will review the impact of technological obsolescence on the Tamil digital documents created in the past 25 years and compare it to the impact on English language digital documents. We will also study the preservation and archival strategies that are evolving in the rest of the world to address the threats to the stability of digital documents. Drawing inspiration from the seminal paper by Paul Convey [3], we hope that this paper will stimulate in-depth discussions among the technical specialists and the broader audiences interested in preserving not only the historical and cultural heritage of Tamils but also the more mundane public records that the government and the citizens depend on across generations.

2. What is the problem? Digital Dilemma!

All current digital storage media are ephemeral. As Conway’s graph demonstrates the millennia-old Indus valley signs and the ancient Tamil inscriptions, while fragile, are still legible [2]. Even the palm manuscripts have managed to survive long enough to transfer information to the future generations.

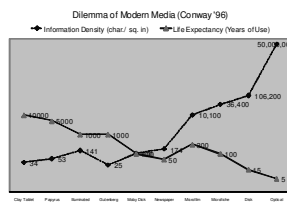


Figure 1. Information Density vs. Life Expectancy [2]

Books printed on acidic paper have lasted more than a century. While the newer recording media allow us to pack enormous information in much smaller area, the shelf-lives of the media, the machines that read them and the applications that render them are all very short (see Figure 1).

2.1 Threats to the Digital Storage

Researchers have identified several threats to the longevity, integrity, access and quality of the digital information storage. Some of these are Media decay and failure, Bit rot, Outdated media, Massive storage failures, Network failure, Access component obsolescence, Outdated formats, Applications and systems failure, Natural Disasters, Human and Software errors, External attacks, Insider attacks, Economic failure, Organizational failure, Politics and Censorship.

In the Tamil context, we can cite some prominent examples. Ambalam.com, a webzine that was launched before Tamil Internet 2000, with a pre-eminent Tamil writer “Sujatha” Rangarajan at the helm, is no longer online. The TamilNation.org site, a veritable encyclopedia of Tamil themed collections shut down once, was resurrected later only to shut down recently for reasons unknown (see Figure 2).



Figure 2. A Major Tamil website disappears

The status of the archives of Tamil digital data from the late 1980s to early 1990s is unknown. The data formats of the old word processors prominent in the MS-DOS and early Windows platforms are not supported by the vendors. In some cases the vendors are no longer around.

The mail archives of one of the first public mailing lists in Tamil (tamil@tamil.net) do not have its oldest archives (starting from October 1996 to September 1998) online. If a researcher were to look for the e-mail where Kōmakaṇ (aka Rajakumaran, the editor of the Malaysian Tamil weekly “Nayanam”) proposed the word ṇaiyam (இணையம்), it will be difficult to get an authentic version of that e-mail from the tamil.net archives.

The encoding and format changes have impacted Tamil digital information in a major way as well. Several early Tamil mailing lists such as tamil.net, agathiyar@yahoogroups.com, tamil-ulagam@yahoogroups.com, had to switch Tamil encoding from Anjal to TSCII and then recently to Unicode. The archives were however left in their original encoding. ForumHub.com, a pioneer in the web based, public bulletin board successfully migrated from TSCII encoding to Unicode including its archives. Similarly, the webzine Thinnai.com switched from TAB encoding to Unicode successfully including its archives. Project Madurai started with TSCII encoding and it now has files both in TSCII and Unicode encodings. Major encoding conversions seldom go without errors. Initial conversions from

TSCII to Unicode at the Project Madurai site had some minor errors (orphaned characters with dotted circle) which appear to have been cleaned up after user feedback.

The founder of the Agathiyar e-mail group, Dr. Jayabharathi, a Tamil scholar and an author from Malaysia, chose to convert all of the Agathiyar postings in the YahooGroups to Unicode and host them separately at his personal site treasurehouseofagathiyar.net. However, he lost most of the articles that he had hosted at various URLs at the geocities.com site when Yahoo shut down Geocities. Several of the pioneering Tamil mailing lists including Agathiyar were once hosted at the hosting site coollist.com and lost their archives when coollist.com folded. The hosting site eGroups.com was acquired by Yahoo and redesigned as YahooGroups.com. Without Yahoo's rescue, the pioneering Tamil mailings all would have lost all of their archives had eGroups.com folded like coollist.

Of equal concern is the information deluge [3]. With the advent of Twitter and Facebook, there is an explosion of public digital content that is rich with information. However, our ability to create digital information far outstrips our capacity and infrastructure to store, manage and preserve it over time.

2.2 Why is preservation important?

Tamil scholars know very well the importance of preservation of manuscripts and inscriptions in the service of cultural heritage. In the 19th century, critical works of Tamil literature were saved from certain destruction just in time by U. Vē. Cāmiṇāṭaiyer and Ci. Vai. Tāmōtaraṇār. And yet, they were unable to recover some major works such as Vaḷaiyāpati and Kuṇṭalakēci.

The examples cited for preserving the Tamil digital documents are not unique. This experience is common to English language based sites as well. The shutdown of GeoCities.com did bring down a major collection of articles in English. The U.S. Library of Congress reports that 44% of the sites available on the Internet in 1998 had vanished one year later. Another study says that 27 months after publication, up to 13% of online cited sources are irretrievable [4] leading to link rot, the process by which the collection of links on a website gradually point to web pages that have become permanently unavailable. Some of the U.S. Census data are reported to be inaccessible due to rapid obsolescence of hardware and software formats. The data recorded by two Viking space probes sent to Mars on magnetic tapes got corrupted despite being kept in a climate controlled environment. Individual consumers are not immune either. Their personal collection of digital photos, documents, e-mails on their floppy disks, portable drives, backup tapes, etc., are equally vulnerable. A lot of consumers have audio cassettes, video tapes in BetaMax and VHS formats that are becoming unusable as machines needed to play them are no longer available.

As the examples cited demonstrate even in the short duration of Tamil digital information's existence, it has experienced the challenges forecast for global digital information storage. This is only going to get more challenging when e-Governance projects are executed in Tamil and millions of Tamil users access e-Governance kiosks to create Tamil documents. The need for planning digital preservation is great.

3. Long-term Digital Preservation

What is long-term digital preservation? And how long is "long-term"? "Long-Term Digital Preservation (LTDP) is a means of keeping digital information such that the same information can be used at some point in the future in spite of obsolescence of everything: hardware, software, processes, format, people, etc." [5] "Long-term" is taken to be the case where it is impossible for a writer to converse with an

eventual reader and impossible for the reader to clarify uncertainties by asking the writer or the writer's contemporaries. [6] In other words, if you write a will bequeathing your wealth to your grandchild and save it in a CD, your grandchild should be able to read what you wrote and prove it to others when you are no more. Your grandchild should do so even if the CD drive, the PC, the Windows operating system, CD driver, the software that reads the contents of the CD and the computer monitor all familiar to you are no longer available and even the CD media itself may have failed. It is not easy. It is not cheap. It will require a lot of organization. And in the end, in spite of best efforts it may not be possible to satisfy all the requirements.

3.1 Challenges of Digital Preservation

Unlike analog data like an old black and white photo, digital data does not degrade gracefully. It requires a specific environment, if not an exact one, to be accessible. Digital files on a medium need specific device drivers to access the medium, specific operating systems to understand the file system and launch a specific application program that knows how to interpret the content of the file. These further depend on specific hardware. If any of these fail, the digital data are not accessible. Digital Preservation aims at packaging the digital objects such that their authenticity is provable and the data are accessible as long as there is a need to read that object. Three main methods of digital preservation are technology preservation, technology emulation and information migration.

3.1.1 Technology Preservation

This is the "computer museum" solution that assumes that the computer/equipment, operating systems, original application software, media drives, etc., all can be preserved, and maintained and that the media is readable. This is the method that was used to recover the pictures taken in 1966 by NASA's robotic probe Lunar Orbiter 1 and stored on 2500 tapes that needed a specialty FR-900 Ampex tape drive, very few of which were made and sold at the cost of \$330,000. It took considerable effort by the very few technicians who had the special skills to work with the reader, a lot of salvaged parts and some good luck. Since there is no commercial or practical interest in such equipment, it is impossible to keep such obsolete equipment functional forever. It worked in this case because of the rarity of the digital object that people were trying to retrieve. [7]

3.1.2 Technology Emulation

Emulation uses a special type of software, called an emulator or virtual machine, to translate instructions from original software to execute on new platforms. This is a fairly well understood technology and practical applications do exist. For example, the popular software VMware offers legacy emulation mode to read DVD-ROM or CD-ROM drives. However, it is difficult to predict how many generations of such emulation will be possible. And for very long-term preservation of the order of several decades to a century or more, technology emulation is yet to be proven.

3.1.3 Information Migration

Information migration requires periodic transfer of digital objects from one system software configuration to another or from one generation of computer technology to a subsequent generation with no loss in content or context and little or no loss in structure. This may require that encrypted digital objects be decrypted. Careful processes need to be evolved to guarantee object authenticity and integrity with this method.

Some of the early creators of Tamil digital objects used a migration method to transfer digital data from MS-DOS based applications and file formats across multiple generations of Windows operating system releases and several versions of applications. Even on the web, users migrated data from one provider to another, as well as from one encoding to another using the conversion tools available. As the Government of Tamil Nadu embarks on the e-Governance program, its offices across the state will be migrating their existing digital data files from various encodings such as TAB, TAM, Baamini, Vanavil etc., to Unicode. This migration of information avoids obsolescence not only of the physical storage medium, but of the encoding and format of the data.

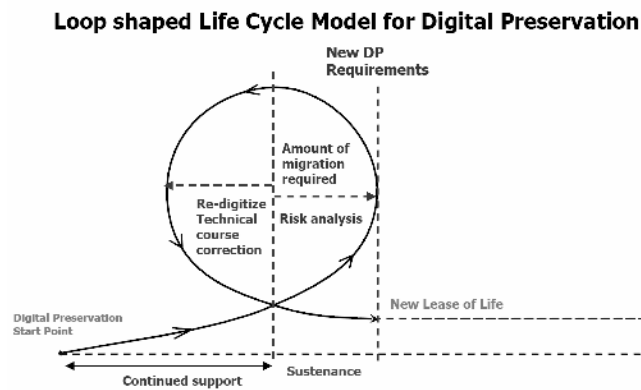


Figure 3. CDAC-Katre Loop shaped life cycle model

In a sense, the Tamil scribes who have been migrating palm manuscripts in Tamil across the centuries have been doing precisely that and have been largely successful in transmitting heritage information to later generations. While this is still error-prone across very long time scale, it is a proven method. This requires that digital objects be constantly migrated to use new storage media, software and computers whenever a current system is in danger of obsolescence.

Loop shaped Life Cycle Model for Digital Preservation

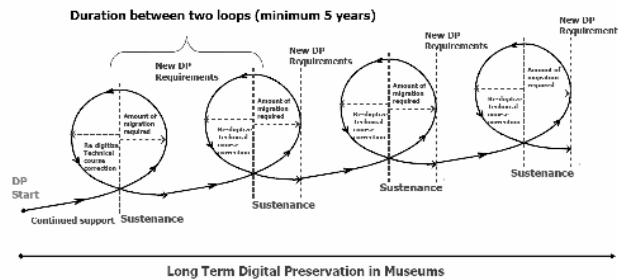


Figure 4. CDAC-Katre Multi-Loop life cycle model

Dr. Dinesh Katre describes a Loop shaped Life Cycle Model for Digital preservation and how the cycle repeats as obsolescence kicks in. (See Figure 3 and Figure 4) [9]

3.1.4 Distributed Digital Preservation

One of the ways in which the authenticity and integrity of digital objects can be preserved is by Replication. Multiple copies also improve the longevity of digital objects if multiple storage locations are used as well. This technique is also known as Distributed Digital Preservation (DDP). This requires extreme discipline, high organization and dedication while the risk of losing authenticity and integrity is high unless one is careful. Historically, this method has been used successfully by religious orders, monks and traditional scholarly communities dedicated to transmitting specialized knowledge across generations. It is noteworthy that a lot of the Tamil palm manuscripts were found in Saiva and Vaishnava mutts, traditional medicinal families, and traditional Tamil scholar families.

A classic example of DDP is the LOCKSS (Lots of Copies Keeps Stuff Safe)[8] technology used by a consortium of publishers, librarians, and learned societies to support a community-managed failsafe repository for scholarly content (<http://lockss.stanford.edu/>).

3.1.5 Metadata and Digital Preservation

Metadata (data about data) is critical to the authenticity and integrity of digital preservation. Metadata is concerned with content, context, and structure of the digital data. Content is intrinsic to the digital information and refers to what the object contains. Context indicates who, what, why, where, how and other data associated with the object's creation and is extrinsic to the object. Structure describes the associations within or among individual objects and can be either intrinsic or extrinsic.

The key to successful implementation of a digital preservation effort are metadata.

There are three different types of metadata, all essential to ensure the usability and preservation of the collection over time.

Descriptive Metadata: convey some sense of intellectual content and context.

Structural Metadata: describe the attributes of an object, such as size, electronic format etc.

Administrative Metadata: information related to rights management, creation date of digital resource, hardware configuration etc. [10]

4. Current Standards and Best Practices

There are several standards that address digital preservation that are of interest to us. The Open Archival Information System (OAIS) was initiated by Consultative Committee for Space Data System (CCSDS) in 2002 and became an international standard with ISO 14721 : 2003. This is the reference model for several other standards. LOCKSS and CLOCKSS comply with the OASIS model and are popular with digital librarians and research publications. The Victorian Electronic Record Strategy (VERS) is a practical specification that addresses the immediate problems of handling large digital documents in popular formats. The National Electronic Records Archive (NARA) consolidates several models with an aim to preserve government records. There are several digital preservation initiatives in India as well, with Digital Library of India (DLI) being the one that digitizes books in Indian languages. There doesn't appear to be any special program anywhere in the world that addresses the unique challenges of Tamil digital documents.

5. Issues of Interest for Tamil documents

Tamil digital documents are constantly being created not only in mainland Tamil Nadu but also in Eezham and the vast Tamil diaspora. While the diaspora largely shares common interests in classical Tamil, when it comes to modern culture, literature, politics, cinema, lifestyle etc., it is quite diverse and differs significantly from the mainland Tamils. Not all of the political opinions of the diaspora are appreciated or even tolerated everywhere. As a result, unpalatable opinions or culturally variant expressions of the diaspora are likely to be ignored if not outright censored by a conservative digital librarian. Selection of the digital content to be preserved is a matter of judgment and since it is expensive to preserve digital content for a long time, it would be tempting to reject the content that one disagrees with regardless of how representative such content may be of contemporary Tamil culture. The digital librarians need only to look at the anthologists that put together the Caṅkam anthologies that have stood the test of time as a guide for selection.

There are some interesting lessons to learn from the fact that the Indus Civilization seals are yet to be satisfactorily decoded and how the Rosetta Stone helped decode Egyptian hieroglyphic writings. The Rosetta Stone had carved text made up of three translations of a single passage: two in Egyptian language scripts (hieroglyphic and Demotic) and one in classical Greek. This is a classic example where Arthur C. Clarke's Rule of Three worked. As interpreted by Mike Shea (mikeshea.net) this rule becomes "preserve in three formats on three media types in three places". Extending this to Tamil documents, one could save important Tamil documents in three different encodings (Unicode, TACE-16 and ISO-15919), in three different formats (PDF/A, XML, HTML) in three different places (Tamil Nadu, Malaysia and Canada). Since the Unicode Tamil block could be used to encode texts in Sanskrit and Saurastri and ISO-15919 could be used to encode all Indic languages, only TACE-16 encoding exclusively encodes Tamil. The subscript/superscript modifiers to indicate varga sounds of other Indic languages do not mix well with TACE-16. So, this could serve as the Rosetta stone for a future researcher if there is any doubt as to the identity of the language of a Tamil digital document. Since the Māṅkuḷam Tamil brahmi inscriptions were thought to be a prakrit inscription for a long time, this is not an idle speculation. The metadata also should be used to clearly identify the language and encodings used to avoid any misidentification. Since such triple encoding preservation is expensive this should be considered only for important documents.

Conclusions

The reality is that solution to the LTDP problem is a work in progress. Some acceptable solutions exist for static data such as documents where the current process is to store in standard encapsulated formats such as PDF/A, wrap with metadata as defined by OAIS and migrate when necessary to address media obsolescence and technology following the CDAC-Katre multi-loop life cycle model. Even here, critics are unsure of the selection of PDF/A for an encapsulated model, preferring HTML or XML. More research is being done to understand the preservation requirements for dynamic data.

We hope that this paper sparks an interest in the long-term preservation and archival challenges to preserve Tamil digital information. The Tamil Virtual University could be one of the best Digital Librarians for preserving all Tamil and Tamil related digital documents and could be one of the nodes of a distributed archival center, perhaps using the LOCKSS technology.

Since large number of Tamil digital documents exist with multiple data formats and encoding, it is hoped that accurate metadata is specified to encapsulate these documents in their original format before migrating them to one standard encoding or three standard encodings (Unicode, TACE-16 and ISO 15919) as discussed earlier.

We hope that this paper would spark an interest in and stimulate in-depth discussions among those interested in very long-term preservation of Tamil digital documents.

References

- [1] Muracu Neṭumārāṇ, 2007. Malēciyat Tamīlarum Tamīlum. International Institute of Tamil Studies. pp 250-264.
- [2] Conway, P. 1996. Preservation in the Digital World , CLIR Reports, Pub 62, 24 pp. ISBN 1-887334-49-1. (<http://www.clir.org/pubs/reports/conway2/index.html>)
- [3] Hey, T. and Trefethen, A., 2002. The Data Deluge: an e-science Perspective. In: Berman, Fran (Ed.) et al, 2003, Grid Computing: Making the Global Infrastructure a Reality, (John Wiley and Sons).
- [4] Dellavalle, R. P. et. al. Information Science: Going, Going, Gone. Science302, no. 5646 (Oct. 31, 2003), 787-8.
- [5] Factor, M. 2010. Long-term Digital Preservation: a View from IBM Research. The Irish Universities Information Services Colloquium, March 2010. http://www.iuisc.ie/2010/Micheal_Factor.ppt
- [6] Gladney, H. M., 2009. Critique of Architectures for Long-Term Digital Preservation. (Draft). <http://eprints.erpanet.org/162/01/LDPCrit.pdf>
- [7] Johnson Jr. J. 2009. NASA's early lunar images, in a new light. The Los Angeles Times, March 22, 2009. <http://articles.latimes.com/2009/mar/22/nation/na-lunar22>
- [8] A Guide to Distributed Digital Preservation. K. Skinner and M. Schultz, Eds. (Atlanta, GA: Educopia Institute, 2010).
- [9] Katre, D. 2008. Imperatives for Survival of Digital Preservation in Indian Museums (A Case Study), National Workshop on Digital Preservation in India, Nov. 7, 2008

இணையத்தில் தமிழ் மின் அகராதிகள்

ஒரு பார்வை

மணியரசன் முனியாண்டி, எம்.ஏ .

விரிவுரையாளர்,

மலேசிய ஆசிரியர் கல்விக்கழகம்,

மலேசியா

maniyarasam.muniandy@gmail.com

இணையத்தில் செயல்படும் தமிழ் மின் அகராதிகளின் பக்கங்களுக்குச் சென்று, அவை வழங்கும் சேவைகளையும், சிறப்புகளையும் தொகுத்து வழங்குகிறார் கட்டுரையாளர். மின் உலகில் தமிழ் மொழி சார்ந்த அகராதிகள் ஐம்பதிற்கும் மேல் உள்ளன. அவற்றின் சேவைகளும் மிகத் தரமான முறையில் அமைந்துள்ளன என்றால் மிகையாகாது. மின்னியல் அகராதி, மின்னணுவியல் அகராதி, கணினி கலைச்சொல் அகராதி எனப் பற்றபல அகராதிகள் சேவையில் இருந்தாலும், தமிழ் மொழிபெயர்ப்பு என்று ஆராயும்போது, ஆங்கே ஏமாற்றம் வரவே செய்கின்றது. பிற மொழிகளுக்கு உள்ள வசதி போல், தமிழ் மொழிக்கு மொழிபெயர்ப்பு வசதிகள் இன்னும் ஏற்படவில்லை என்று உறுதியாகக் கூறினாலும், அதற்கான ஆயத்தப் பணிகள் மேற்கொள்ளப்பட்டு வருகின்றன என்ற கூகள்(google) செய்தி நிம்மதி தருகின்றது. சுருங்கக்கூறின், தமிழ் மின் அகராதிகள் தமிழுக்கு ஊக்கத்தினை மட்டும் அல்லாமல் ஆக்கத்தினையும் ஏற்படுத்தும் என்று நம்பலாம்.

முன்னுரை:

இணைய அரங்கினில் ஆங்கில மொழிக்கு ஈடாக நடை பயிலும் மொழிகளுள் தமிழ்மொழியும் ஒன்று. கணினியின் துணைகொண்டு ஆங்கில மொழியில் முடிக்கவல்ல அனைத்து மின் பணிகளையும் தமிழில் முடிக்கலாம் என்ற நிலை ஏற்பட்டு நெடுநாட்கள் கடந்து விட்டன. முன்பு, தமிழ் வலையகங்களில் உலா வருவதற்குத் தமிழ் எழுத்துருக்களைப் பதிவிறக்கம் (Download) செய்யவேண்டிய நிலைகள் இருந்தன. அச்சுழல் கடந்து, இன்று தானிறங்கி எழுத்துமுறை (Dynamic Font) நடைமுறையில் உள்ளது. தமிழில் தட்டச்சு செய்தல், அதனை வடிவமைத்துத் தொகுத்தல், வரைகலை மேற்கொள்ளல், கானொளி அமைத்தல், மின்னஞ்சல் அனுப்புதல் மற்றும் பெறுதல், இணையத்தளங்கள் மற்றும் வலைப்பூக்கள் வடிவமைதல், தேடுதளங்களை உருவாக்குதல், மின்னிதழ்கள் அல்லது இணைய இதழ்கள் வெளியிடுதல், இணைய வானொலி தொலைக்காட்சி நடத்துதல் முதலான அனைத்துப் பணிகளையும் தமிழிலேயே செய்துகொள்ளமுடியும் என்ற நிலையைச் செம்மொழியாம் நம் தமிழ் அடைந்துவிட்டதைக் காண முடிகின்றது.

இன்றைய அவசர உலகில், ஏடு வடிவில் இருக்கும் அகராதியைவிட, மின் வடிவில் இருக்கும் அகராதி பேருதவியாக இருக்கின்றது. இணையம் வழியாகச் செயல்படும் தமிழ் மின் அகராதிகளைத் தமிழ் மக்களின் பார்வைக்குக் கொண்டு வரவேண்டும் என்ற நோக்கில் இக்கட்டுரை வரையப்படுகின்றது. நொடிப் பொழுதில் தமிழ் சொற்களுக்கு ஏற்ற ஆங்கிலச் சொற்களையும், ஆங்கிலம் முதலிய இதர அயல் மொழிகளுக்கு ஏற்ற தமிழ்ச் சொற்களையும் அறிந்து கொள்ள மின் அகராதிகள் நன்கு உதவுகின்றன

என்பதில் ஐயமில்லை. கற்றல் கற்பித்தலில் அகராதியின் பங்கு முதன்மையானதாகத் திகழ்கின்றது. இணைய எழுத்துப் படிவங்கள் சிறப்பாகவும் இலக்கண அழகோடும் மிளிர், மின் அகராதிகள் அதிகம் உதவுகின்றன என்பதில் எந்தவொரு ஐயப்படுமில்லை. சிறப்பானதொரு தமிழகராதி, இணையத்தை அறிவுக் களஞ்சியமாக எண்ணும் நிகழ்கால மாணவர் சமுதாயத்திற்குச் சிறந்த வழிகாட்டியாக அமையும் என்பதும் திண்ணமாகும்.

தமிழ் அகராதி சேவை வழங்கும் இணையத் தளங்கள்

1. <http://www.thozhilnutpam.com/chollagaraathi.htm> என்ற இணைய முகவரியில் இயங்கும் 'தமிழ் மின்னியல் மின்னணுவியல் கணினியியல் சொல்லகராதி' சிறந்த பல கலைச்சொற்களைத் தரும் மின் அகராதிகளுள் ஒன்றாகும். மேற்குறிப்பிட்ட இணையத்தளத்தின் உதவியினால், மின்னியல், மின்னணுவியல் மற்றும் கணினியியல் ஆகிய துறைகளில் பயிலும் மாணவர்கள் அல்லது அத்துறைகளிற் சார்ந்து எழுத்துப் படைப்புகளை வெளியிடும் எழுத்தாளர்களும் அவரவர்களுக்கு வேண்டிய சரியான கலைச்சொல்லினைப் பெற்றுக்கொள்ளலாம்.

மின்னியல்(லெட்ரிக்), மின்னணுவியல் (இலட்ரோனிக்) மற்றும் கணினியியல் (கம்பியூட்டர்) போன்ற துறைசார் கலைச்சொற்களுக்கு இங்குப் பஞ்சமே இல்லை எனலாம். இவ்வகராதியினைச் செயல்படுத்தும் தமிழ் தொண்டாளர்கள் தமிழ் மக்களிடமிருந்து பற்பல புதிய சொற்களை எதிர்பார்ப்பதாகக் குறிப்பிடுகின்றனர். தமிழ் மின்னியல் மின்னணுவியல் கணினியியல் சொல்லகராதி (TAMIL ELECTRICAL, ELECTRONICS & COMPUTER ENGINEERING GLOSSARY) மற்றும் தமிழ் மின்னியல் மின்னணுவியல் கணினியியல் சொல்லகராதி (TAMIL ELECTRICAL, ELECTRONICS & COMPUTER ENGINEERING GLOSSARY) ஆகிய இரு மின் அகராதிகளும் நமது தேவையைப் பூர்த்தி செய்கின்றன.



மேற்படியுள்ள குறிப்பெட்டியை குறியிட்டு APPLY அமுக்குக.

விரிவாக்கப்பட்ட கோப்புகள் உள்ள கோப்புறையை தேர்வு செய்க.

OK அமுக்குக.


(விரிவாக்கப்பட்ட கோப்புறையில் இல்லாத கோப்புகளை கேட்கும்போது CANCEL, YES அமுக்கி தொடர்க.)

கிடைசியில் RESTARTற்காக கேட்கும்போது, RESTART செய்க.

2. <http://www.tamilvu.org/library/o33/html/o3300001.htm>

இணையத்தில் தமிழ் அகராதி சேவையினை வழங்கிச் சிறப்புடன் செயல்படும் மின் அகராதிளுள் குறிப்பிடத்தக்கது 'தமிழ் இணையப் பல்கலைக்கழகத்தின்' மின் அகராதியாகும். இந்தத் தமிழ் மின் அகராதியில் சமுதாயவியல், மருத்துவவியல், கால்நடை மருத்துவவியல், உயிரிய தொழில் நுட்பவியல், கலை மானுடவியல், தகவல் தொழில் நுட்பவியல், தொழில் நுற்பவியல், பொறியியல், தொழில்நுற்பவியல், வேளாண்மைப் பொறியியல், அறிவியல், சட்டவியல், மற்றும் மனையியல் போன்ற துறைகளுக்கான கலைச்சொற்களுக்குப் பொருள் சுட்டப்பட்டுள்ளன. இணையப் பயனாளர்கள் இச்சேவையை நன்கு பயன்படுத்திக்கொள்ளலாம்.

3. <http://www.tcwords.com/>

இந்த  இணைய அகராதியின் வழியாகத் தமிழ்ச்சொற்களுக்கு ஏற்ற, சரியான ஆங்கிலச் சொற்களையும் அறிந்து பயனடையலாம். தமிழ் உரைநடைப் பகுதியை ஆங்கில் மொழியாக்கம் செய்யும் ஆர்வலர்கள் அவ்வகப்பக்கதின் உதவியை நாடிப் பயனடையலாம்.

4. TAMIL COMPUTING WORDS , TAMIL VIRTUAL UNIVERSITY DICTIONARIES, ANNA UNIVERSITY ENGLISH-TAMIL COMPUTER DICTIONARY.

அண்ணா பல்கலைக்கழகத்தின் சார்பில் வளர்தமிழ் மன்றம் வெளியீடு செய்த 'கணிப்பொறிக் கலைச்சொல் அகராதி' இணையத்தாருக்கு மிகுந்த பயனை அளிக்கின்றது. பன்னிரு தமிழ் அறிஞர்கள் ஒன்றிணைந்து ஆங்கிலத்திலிருந்து தமிழுக்கு மொழிபெயர்த்து வழங்கிய அச்சொத்து, இந்த நூற்றாண்டின் அரிய படைப்பாகும். மேற்குறிப்பிட்ட மூன்று இணையத்தள அகராதிகளையும் பின் குறிப்பிடப்படும் விவரங்களில் காணலாம்

5. <http://www.tamildict.com/deutsch.php?menu=new&action=new>

tamildict.com

English-Tamil-German Dictionary

இந்த மின் அகராதியும் தமிழ்ச் சேவையில் முதன்மை வகிக்கின்றது. ஆங்கிலம் - தமிழ்- ஜெர்மன் ஆகிய மும்மொழி அகராதி என்ற சிறப்பினை இவ்வகராதி பெறுகின்றது. புதிய தமிழ்ச்சொற்கள் வேண்டும் என்ற வேண்டுகோளையும் இவ்வகப்பக்கம் விடுத்துள்ளது.

6. Digital Dictionaries of South Asia

டிஜிட்டல் டிசுஷனரில் ஆப் சவுத் ஆசியா, என்ற பெயரிலான மின் அகராதி இணையப் பயனாளர்களுக்கு வேண்டிய அனைத்து வேலைகளையும் நொடிப்பொழுதில் தீர்க்கும் வல்லமை பொருந்தியதாக இருக்கின்றது. 'அன்பு' என்ற சொல்லுக்கு அளிக்கப்பட்ட சொற்பொருள் விளக்கம் அடியில் கொடுக்கப்பட்டுள்ளது. இணையப் பயனாளர் வசதிக்காக அதன் அகப்பக்கமும் கொடுக்கப்பட்டுள்ளது.

1. அளி (p. 40) [a| i] , s. gift, கொடை; 2. favour, அருள்;
3. desire, ஆசை; 4. love. அன்பு; 5. civility, உபசாரம்; 6. poverty, வறுமை;
7. unripe fruit, காய். அளியன், one who is kind; 2. a paupe

நாம் தேடிச்செல்லும் சொல்லுக்குரிய பொருளினைக் கனநேரத்தில் தந்து சிக்கல் தீர்க்கும் அரிய நண்பன்தான் Digital Dictionaries of South Asia. இன்றைய மாணவர்கள், குறிப்பாகக் கல்லூரி மற்றும் பல்கலைக்கழகங்களில் தமிழ் பயிலும் இளங்களை, முதுகலை மாணவர்கள் மேற்கோளுக்கான நாடும் ஒப்பற்ற இணையத்தலம் இதுவாகும்.

7. Google English - Tamil dictionary and Google Tamil - English dictionary

கூகில் தமிழ் ஆங்கில அகராதி தனிச்சிறப்பு வாய்ந்ததாக இருக்கின்றது. வழக்கமான நிலையிலிருந்து, சொற்களுக்கேற்ற பொருளை மட்டும் காட்டாமல், பட விளக்கத்தினையும் தருகின்றது இந்த மின் அகராதி. தொடக்க நிலையில் தமிழ் கற்பவர்களுக்குச் சிறந்த துணைவனாக இவ்வகராதி திகழ்கிறது என்வதில் ஐயமில்லை.

About the Dictionary	Front Matter	Back to DDSA
Search Help		Font Help
Look up an entry word	<input style="width: 50px;" type="text" value="? ???"/>	(Example: <u>அகத்தியன்</u> or encomium)
<input style="width: 50px;" type="button" value="தேடு"/> <input style="width: 50px;" type="button" value="தேடு தமிழில்"/>		

8. Online Tamil Dictionary

மேலும், 'ஆன்லைன்' அகராதிகள் என்றும் பல மின் அகராதிகள் இணையதில் செயல்படுகின்றன. அவற்றுள் சில கட்டணங்கள் வசூலிக்கும் பொருளாதார நோக்குடையவை என்பது ஆராய்ந்தால் புலனாகின்றது. அவையாவன:

Online Tamil Dictionary, - www.tamildict.com/

Tamil dictionary, - www.tamil.net/learn-tamil/tamildic.html

English to Sinhala and Tamil Online Dictionary from Sri Lanka

www.lanka.info/dictionary/EnglishToSinhala.jsp

Online Dictionaries - Tamil Online

Dictionaries www.multilingualbooks.com > ... > Online Dictionaries

Tamil English

dictionary and English Tamil dictionary - FREELANG www.freelang.net/dictionary/tamil.php

Tamil English and English Tamil Dictionary Free Online Translation www.ats-group.net › Languages › Dictionaries

Tamil and English dictionary - dsal.uchicago.edu/dictionaries/fabricius/ -

Dravidian Languages www.yourdictionary.com › Languages

Sanskrit, Tamil and Pahlavi Dictionaries - webapps.uni-koeln.de/tamil/

alphaDictionary * Free Tamil Dictionary - Free Tamil Grammar www.alphadictionary.com › Languages

› Dravidian -


bottle in Tamil - Google Dictionary Page 1 of 2

[Web](#) [Images](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼ maniyarasan.muniandy@gmail.com | [Web History](#) | [My Account](#) | [Sign out](#)

[Go to Google Home](#)

Dictionary

Found in dictionary: **English > Tamil**.

☆ **bottle** 

1. புட்டி

Related phrases

aspirator bottle
உறிஞ்சுக் குப்பி

magnetic bottle
காந்தக் குடுவை

bottle resonator
போதற்பரிவுக்கருவி

bottle feeding
புட்டுப் பாலூட்டல்

blue bottle
நீல ஈ

bottle neck
இடர்ப்பாடு

ink-bottle
மைக்கூடு

bottle-shaped jack
போதல்வடிவத்தூக்கி

green bottle fly
பச்சை ஈ

bottle shape jack
உருளைவடிவ பளுதூக்கி

Synonyms
noun: [flask](#)

Usage examples

"Since I had to leave the **bottle** behind, I asked that it be destroyed. At no point did I request preferential treatment; it's not in my nature," Blackburn said in his [statement](#).
Mar 19, 2010 - [Jean-Pierre Blackburn](#) - [Toronto Star](#)

"I have repeatedly noted that we haven't turned any corners," Petraeus said. "We haven't seen any lights at the end of the tunnel. The champagne **bottle** has been pushed to the back of the refrigerator. And the progress, while real, is fragile..
Apr 8, 2008 - [David Petraeus](#) - [Boston Globe \(registration\)](#)

Winslet, who has been nominated five previous times, said: "I'd be lying if I said I haven't made a version of this speech before. I think I was probably 8 years old and staring into the bathroom mirror, and this would have been a shampoo **bottle**....
Feb 22, 2009 - [Kate Winslet](#) - [ABC News](#)

Web definitions

- a glass or plastic vessel used for storing drinks or other liquids; typically cylindrical without handles and with a narrow neck that can be ...

<http://www.google.com/dictionary?langpair=ta%7Cen&q=bottle&hl=en&aq=f> 4/10/2010

9. வளரும் நிலையில் உள்ள பற்பல தமிழ் மி அகராதிகளுள் ஒன்று:-

'காதல்' என்ற சொல்லுக்குப் பொருள் தேடப்பட்டபோது, அவ்வாறான சொல்லே இல்லை என்ற பதில் தந்த மின் அகராதியும் இணையத்தில் உண்டு. பொது நலன் கருதியும், அதன் நடத்துநர் நலன் பொருட்டும், அதன் இணைய முகவரி கொடுக்கவில்லை. இவ்வாறான மின் அகராதிகள் இன்னும் சிறப்பான சேவையினை வழங்கக் காலம் கனிய வேண்டும்.

Utilities

Transliteration

Dictionary

Select Language

Home

Help

English Word There is no result found

10. இலவயமாக செயல்படும் மேலும் சில தமிழ் மின் அகராதிகள்:-

tamil online	Tamil Wiktionary
DSAL dictionaries	J. P. Fabricius's
Tamil and English dictionary	Tamil Moli Akarathi. N.
Kathiraiver Pillai's Tamil Moli Akarathi: Tamil-Tamil dictionary = Na. Kathiraiver Pillayin Tamil Moliyakarati: Tamil-Tamil akarathi.	A core vocabulary for Tamil
Tamil lexicon.	A comprehensive
Tamil and English dictionary of high and low Tamil	Pals e-dictionary
Tamil Virtual University Dictionaries	கரியாவின்
தற்காலத் தமிழ் அகராதி dictionary and Google Tamil – English dictionary	Google English – Tamil

தமிழ் மொழிபெயர்ப்புச் சேவை வழங்கும் இணையத் தளங்கள்

1. Google Transliteration IME

New! Download Google Transliteration IME Type a word in English and press SPACE to transliterate. Press CTRL+G (# +G on Mac) to switch between English and the selected language.

Dismiss

©2009 Google - Font Guide - Discuss - Help - Google Home

Define

Translate

?? ?? ???????

English

?? ??

From

Tamil

To

English

swap

கனவு

1. dream
2. morpheus

Powered by Google Dictionary

'கூகிள்'(google) தயவினில் தோன்றிய புத்தம்புதிய வரவு 'கூலில் திரான்ஸ்லிதரேஷன்' அகராதி யாகும். இந்த ஒலிபெயர்ப்பு அகராதியின் துணைகொண்டு ஆங்கிலச் சொல்லை ஒலிபெயர்த்துக் கொள்ளலாம். இணையத்தில் இயங்கும் மின் அகராதிகளுள் இதுவும் சிறப்பானது என்பது குறிப்பிடத் தக்கது.

2. <http://translate.google.com/toolkit/list#translations/active>

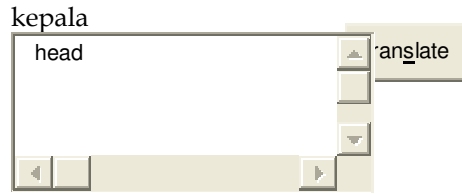
முழுமையான தமிழ் மொழிபெயர்ப்புக்கு Google Translator Toolkit என்ற மென்பொருளின் முகப்புப் பகுதிக்குச் செல்ல வேண்டும். <http://translate.google.com/toolkit/list#translations/active> என்ற தளத்திற்கு சென்று காண்க. இருப்பினும், தரமான முறையில் மொழிபெயர்ப்புச் செய்யவேண்டும் என்றால் இயலாத காரியமாகும். இதன் காரணம், தமிழ் 'கூகில்' பகுதியில் தரமான தமிழ் அகராதி இன்னும் அமையவில்லை என்பதுதான்.

3. http://www.stars21.com/translator/english_to_indonesian.html

ஸ்டார்ஸ்21 என்ற இணையத்தள முகவரியில் செயல்படும் மின் அகராதி மலாய் மொழியிலிருந்து ஆங்கில



மொழிக்கு மொழிபெயர்க்கின்ற வசதியைத் தருகின்றது. செம்மொழியாகிவிட்ட தமிழ் மொழிக்கு இங்கு இடமில்லை என்பது வருத்தம் அளிக்கும் செய்தியாகும். ஸ்டார்ஸ் 21 என்ற அகப்பக்கத்தின் முகப்பு, கீழே வழங்கப்பட்டுள்ளது.

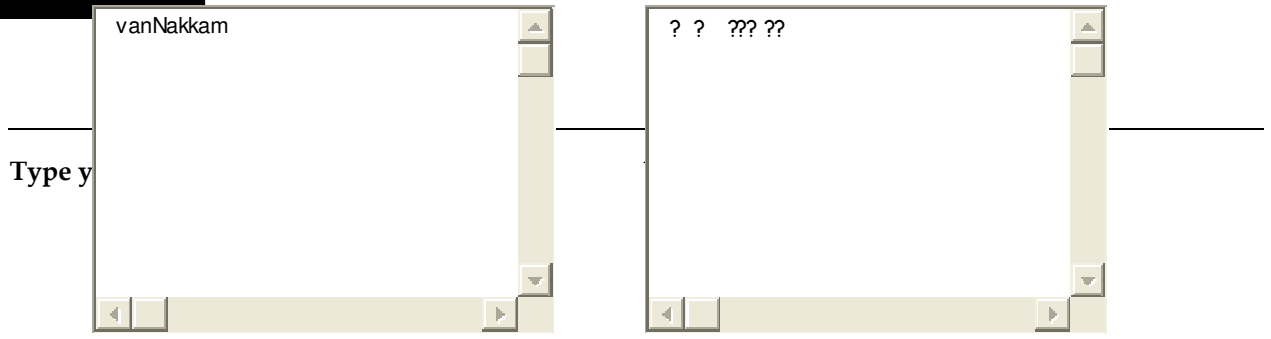


head

4. <http://www.tamilcube.com/res/tamilpad.html>

மேற்கண்ட இணையத்தளத்தில் பொருள் தேட விரும்பினால், பின் குறிப்பிடப்பட்ட வழிமுறைகளைப் பின்பற்றவேண்டும். VANAKKAM என்று ரோமன் முறையில் விசைத்தால், தமிழ் எழுத்தில் 'வணக்கம்' தென்படுகின்றது. தொடக்க நிலையில் தமிழ் கற்போர் இதனைப் பயன்படுத்தி நன்மை அடையலாம்.

(TamilCube's Tamilpad is a free online Tamil language typing software. Tamilpad helps you to type in Unicode Tamil. Just type in English in the left box, by following the Transliteration keyboard mapping given below. The equivalent Unicode Tamil letters appear automatically in the right box. For example, when you type 'vaNakkam' in English and hit the space bar, Tamilpad will convert it directly into Unicode Tamil script as 'வணக்கம்'. You can cut and paste these Unicode Tamil words from Tamilpad into your email messages such as Yahoo or Hotmail or into any editor such as Microsoft Word or Microsoft Excel).



<http://www.tamilcube.com/dictionary>

Enter your English or Tamil word for translation in the search box below and click 'SEARCH'

This field is required.

English -> Tamil
 Tamil -> English
 Number -> Tamil word

தமிழ் மின் அகராதி வரிசைகளில் குறிப்பிட்டுச் சொல்லும் வகையில் அமையும் பல அகராதிகளுள் 'தமிழ் கியூப்' மின் அகராதியும் ஒன்றாகும். தமிழ் மொழிபெயர்ப்புச் செய்ய விரும்பும் ஆர்வலர்கள் இந்த அகப்பக்கத்தினை நாடிப் பயனடையலம்.

Modern Online Tamil dictionary (English-Tamil & Tamil-English)

Browse for basic words in online Tamil dictionary: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Tamilpad - TamilCube's Free Online English to Tamil Transliteration tool

TamilCube's Free Online English to Tamil Dictionary and English to Hindi Dictionary

English to Tamil and Tamil to English professional Translation Service

5. <http://ta.wiktionary.org/wiki/>

112,490 தமிழ்ச்சொற்களுடன் உலா வருகின்றது மேற்குறிப்பிட்ட விக்சனரி தமிழ் மின் அகராதி. கட்டற்ற சொற்களுக்குப் பொருள் வழங்கும் தன்மையை உடைய இவ்வகராதி அனைவரின் கவனத்தையும் ஈர்க்க வல்லது. குறிப்பிட்டதொரு தமிழ் எழுத்தினைச் சொடுக்கினால், அச்சொல்லுக்கான பொருள்கள் விரிகின்றன. தமிழ்ச் சொற்களுக்கு ஏற்ற மற்றொரு சொல்லை அறியவேண்டும் என்றால், அவ்வகப்பக்கத்தினை நாடிப் பயனடையலாம். விக்சனரி அகப்பக்கத்தின் மாதிரி கீழே தரப்பட்டுள்ளது.

6. இதர மின் மொழிபெயர்ப்பு அகப்பக்கங்கள்:

Tamil translation dictionaries - lexicoo...

Online **Tamil** bilingual and multilingual dictionaries (**Tamil** <-> English, French, etc). List updated regularly.

www.lexicool.com/dictionaries_tamil.asp [Found on Google, Bing]

Tamil translation from to English **tamil** ...

All kind of **Tamil translation** and **tamil** language related works are done in time according to your needs.

www.tamiltranslator.com/ [Found on Google, Bing]

English to **Tamil translation**

Your search for English to **Tamil translation** returned 142 results in the following **Translation** of Christian Manuscript from English to **Tamil** ...

www.translatorscafe.com/cafe/PopSearch/English-to-Tamil-tran... [Found on Google, Bing]

Tamil English and English **Tamil** Dictiona...

Free Online Dictionary Welsh English and Free Online **Translation Tamil** English Dictionaries.

www.ats-group.net/dictionaries/dictionary-english-tamil.html [Found on Google, Bing]

Tamil Translation Service

Tamil Translation services from Applied Language Solutions high quality, professional, award winning **Tamil Translation**

www.appliedlanguage.com/languages/tamil_translation.shtml [Found on Bing]

Professional **Tamil translation** service |...

One-stop language resources portal. Professional **Tamil translation** service. Free **Tamil** books, **Tamil** mobile eBooks, **Tamil** test papers.

www.tamilcube.com/ [Found on Google]

English to **Tamil Translation**

English to **Tamil** language **translation** service ... Superb English to **Tamil translation** . In need of a quality **Tamil** to English or English to **Tamil translation**?

www.kwintessential.co.uk/translation/to/tamil.html [Found on Bing]

tamil translation - Google Chrome Help

Feb 10, 2010 ... Among the Indian language versions, is Google Chrome also being targetted only for Hindi-knowing people only? Why is there no **Tamil** option? ...

www.google.com/support/forum/p/Chrome/thread?tid=186d352aa90... [Found on Google]

Tamil translation, English to **Tamil** tran...

wintranslation.com provides professional **Tamil translation** service performed by human translators.

www.wintranslation.com/languages/tamil.html [Found on Bing]

Internet Archive: Free Download: QURAN T...

QURAN **TAMiL** TRANSLATI**ON** NOT ARABIC MP3 This audio is part of the collection: Open Source Audio Artist/Composer: QURAN **TAMiL**. Keywords: ALLAH islam Muhammed ...

www.archive.org/details/tamilquranmp3 [Found on Google]

English **Tamil** dictionary - lexicool.com

English **Tamil** dictionaries ... by language Online dictionaries by subject Lexicool blog Lexicool newsletter **Translation** courses ...

www.lexicool.com/online-dictionary.asp?FSP=A09B31 [Found on Bing]

Tamil Translation - Free with Linguanaut

Our website Linguanaut helps you get free **Tamil translation** from our translator volunteers, like how to say hello, welcome, thank you, other greetings and ...

www.linguanaut.com/translation_tam.htm [Found on Google]

Tamil translation English **Tamil** translat...

Translation India offers **Tamil translation** and English to **Tamil translation** services. Get documents, technical, legal, financial and book **translation** in **Tamil** language by ...

translationindia.com/indian-languages-tamil.html [Found on Bing]

Tamil - Google & BabelFish **translation** i...

Thanks **Tamil**. I also want to add Google **Translation** (site **translation**) of member Apropos. I added the line -Submenu, Google Translate, Google Translate ...

my.opera.com/Tamil/blog/google-babelfish-translation [Found on Google]

Amazon.com: Understanding Muhammad: A Ps...

Understanding Muhammad: A Psychobiography (**Tamil** Edition) (Paperback). ~ Ali Sina

(Author), Mona Malik Mustafa (Translator) ... www.amazon.com/Understanding-Muhammad-Psychobiography-Ali-Si... [Found on Google]

Tamil Translation. Tamil to English Tran...

Professional **Tamil** to English and English to **Tamil translation** and Localization. Rapid response, accurate **translations** available worldwide.

www.worldlingo.com/en/languages/tamil_translation.html [Found on Bing]

2. கைத்தொலைப்பேசியில் மின்னணுவியல் அகராதி

<http://www.tamilcube.com/dictionary/mobile/>

கைத்தொலைப்பேசியில் தமிழ் மின்னணுவியல் அகராதி இணையாதா என்ற ஏக்கம், பின் குறிப்பிடப்பட்ட விளம்பரத்தின்வழி தீரும் என நம்பலாம். 'தமிழ் கியூப்' இந்த சேவையினை வழங்கும் அறிவிப்பினை விட்டிருக்கின்றது. இந்தச் சேவை அனைத்துத் தொலைப்பேசிகளிலும் இடம்பெற்றால், இணைய வசதி இல்லாத இடங்களிலும் தமிழ் செல்லும். பள்ள, கல்லூரி மற்றும் பல்கலைக்கழகங்களில் பயிலும் மாணவர்கள், குறிப்பாக மாணவர்களும், கற்பித்தலை மேற்கொள்பவர்களும் மிகுந்த பயனை அடைவார்கள். தமிழ் கியூப் கைத்தொலைப்பேசி மின்னணுவியல் அகராதியைப் பதிவிறக்கம் செய்யும் முறை காட்டப்பட்டுள்ளது.

Download Mobile Dictionaries

TamilCube welcomes you to the world of Modern Mobile Dictionaries! You can download English-Tamil, English-Malay and English-Hindi dictionary software for mobile to any of your mobile phones with Java support, such as any make and model of mobile phone (cell phone), PDA, iPhone and Blackberry. To view the mobile dictionaries in your mobile phone, just follow the simple steps below:

1. Download the zip file for the dictionary you like, unzip into a folder in your PC.
2. Connect your mobile phone, iPhone or Blackberry to the PC and transfer the unzipped software files to your mobile. (This method is the same as how you download games or applications to your mobile device.) Now you are ready to enjoy referring to the dictionary from your mobile phone, anytime and anywhere!

DOWNLOAD MOBILE DICTIONARY SOFTWARE FOR TRIAL

TamilCube's Modern English-Tamil Mobile Dictionary (Trial)

TamilCube's Modern English-Malay Mobile Dictionary software(Trial)

TamilCube's Modern English-Hindi Mobile Dictionary software(Trial)

- The trial mobile dictionaries contain only the words starting with letter "a".
- To view the Tamil and Hindi mobile dictionary, your mobile phone, iPhone or Blackberry must have Unicode support.
- When searching the dictionary, use lower case English letters.

Translate text, webpages and documents

	en	UTF-8	1	1
--	----	-------	---	---

load a document.

Translate from:

English

Translate into:



- | | | | | |
|------------|----------|----------------|------------|------------|
| Afrikaans | Danish | Greek | Japanese | Polish |
| Albanian | Dutch | Haitian Creole | Korean | Portuguese |
| Arabic | English | Hebrew | Latvian | Romanian |
| Belarusian | Estonian | Hindi | Lithuanian | Russian |
| Bulgarian | Filipino | Hungarian | Macedonian | Serbian |
| Catalan | Finnish | Icelandic | Malay | Slovak |
| Chinese | French | Indonesian | Maltese | Slovenian |
| Croatian | Galician | Irish | Norwegian | Spanish |
| | Italian | Persian | Swahili | |

	mother	en en	
--	--------	-------	--

Submit

மேற்கண்ட கூகிள் (google) மொழிபெயர்ப்புக் கருவியினைப் பாருங்கள். தமிழைப் தவிர ஏனைய மொழிகள் அனைத்தும் உண்டு, தமிழ்தான் இல்லை. இந்த மொழிபெயர்ப்பு கருவியில் தமிழ்மொழி இணக்கப்பட்டால், ஆங்கிலம் போன்ற பிற மொழிகளில் உள்ள ஒரு பத்தியை அப்படியே தமிழுக்கு மாற்றலாம். அக்காலம், என்று கனியுமோ?

முடிவுரை:

சுருங்கக் கூறின், இணையத்தில் செயல்படும் மின் அகராதிகள் தமிழ்மொழியின் மேன்மைக்குப் பெரிய பங்கினை ஆற்றி வருகின்றன. சில முதன்மை அகப்பக்கங்கள் சிறப்பான அகராதிகளைப் பற்றல மொழிகளில் இடம்பெறச் செய்திருக்கின்றன. இது பாராட்டுதற்குரியது என்றாலும், தமிழ் மொழியும் இடம்பெற்றிருந்தால் பயனாக இருந்திருக்கும். தமிழ் மின் அகராதிகள் இணையத்தில் பெருகினால், தமிழ் எழுத்துப்படிவங்களின் தரம் உயரும். பாரதியின் 'பிற நாட்டு சாஸ்திரங்கள் தமிழ் மொழியில் பெயர்க்கப்பட வேண்டும்' என்ற கனவு நனவாகும். தமிழில் புதைந்துள்ள பெருமைகள் உலகின் எட்டுத் திசையெங்கும் மொழிபெயர்ப்பு உதவியால் பல்கிப் பரவும்.

இணையத்தில் தமிழ் நூல்கள்

முனைவர் க.துரையாசன்

இணைப்பேராசிரியர், தமிழ்த் துறை
அரசினர் கலைக் கல்லூரி(தன்னாட்சி)

கும்பகோணம் - 612 001
darasan2005@yahoo.com

முகவுரை

இருபத்தொன்றாம் நூற்றாண்டின் இணையற்ற அறிவியல் சாதனமாம் கணினியில் நூல்களைப் பதிவு செய்து பாதுகாக்கும் பணி உலகெங்கிலும் எல்லா மொழிகளிலும் வெகு வேகமாக நடைபெற்று வருகிறது. அவ்வகையில் தமிழ் மொழியில் உள்ள இலக்கிய, இலக்கண நூல்களைக் கணினியில் பதிவு செய்து சேமித்து வைக்கும் பணி அரசாலும், அரசு சார்புடைய அமைப்புகளாலும், தனியார் பலரின் ஆர்வத்தாலும் நடைபெற்று வருகிறது.

இப்பணியில் தமிழ் இணையப் பல்கலைக்கழகம், இந்திய மொழிகளின் நடுவண் நிறுவனம், மதுரைத் திட்டம், சென்னை நூலகம், விருபா.காம், நூலகம்.நெட், ரோஜா முத்தையா ஆராய்ச்சி நூலகம், விக்ரீடியா முதலானவை குறிப்பிடத்தக்கவை ஆகும். இவை தொடர்பான செய்திகளை எடுத்துரைப்பதாக இக்கட்டுரை அமைகிறது.

தமிழ் இணையப் பல்கலைக்கழகம்

உலகு தழுவிய வாழும் தமிழர்கள், தமிழ் ஆர்வலர்கள், தமிழ் அறிஞர்கள் மற்றும் தமிழ் ஆய்வாளர்களின் தேவைகளை மனத்தில் கொண்டு தமிழ் இணையப் பல்கலைக்கழகம் ஒன்று அமைக்கப்படும் என்று 1999இல் நடைபெற்ற இரண்டாவது உலகத் தமிழ் இணைய மாநாட்டின் நிறைவு விழாவில் தமிழக முதல்வர் முத்தமிழறிஞர் டாக்டர் கலைஞர் அவர்கள் அறிவித்தார்கள். அவ்வறிப்பினை நிறைவேற்றும் வகையில் 17-02-2001இல் இப்பல்கலைக்கழகம் அவரால் தொடங்கப்பட்டது.

இப்பல்கலைக்கழக மின் நூலகத்தில் தற்பொழுது ஏறக்குறைய ஒரு இலட்சத்து ஐம்பதாயிரம் பக்கங்களுக்கு மேலான 300க்கும் மேற்பட்ட தமிழ் நூல்கள் இடம் பெற்றுள்ளன. இவைப் பல்வேறு தேடுதல் வசதிகளைக் கொண்டுள்ளன.

இம் மின்நூலகத்தில் தமிழ் நூல்கள் மட்டுமின்றி தமிழ் நூல்களின் ரோமன் வாவடிவம், அகராதிகள், கலைக் களஞ்சியம், கலைச்சொல் தொகுப்புகள், சுவடிக் காட்சியகங்கள், பண்பாட்டுக் காட்சியகங்கள், பயணியர் தமிழ் முதலியன உள்ளன. மேலும் தமிழர்களின் இறை உணர்வைப் புலப்படுத்துகிற வகையில் சைவ, வைணவக் கோயில்களின் ஒலி, ஒளிக் காட்சிகளும் இடம் பெற்றுள்ளன.

இப்பல்கலைக்கழக இணைய தளத்தில் செவ்வியல் இலக்கியங்களான சங்க இலக்கியங்கள் பழம்பெரும் உரைகளுடன் சிறப்பான முறையில் இடம் பெற்றுள்ளன. இவ்விலக்கியங்களில் தேவையான தகவல்களைத் தேடிப் பெறுகின்ற வகையில் இவை இணையதளப்படுத்தப்பட்டுள்ளன. இவைத் தமிழாய்வுலகிற்குப் பேருதவியாக திகழ்ந்து வருகிறது. திருக்குறளுக்குக் கலைஞர் உரை முதலாக அறுவரின் உரைகள் ஒப்பிட்டுப் படிக்கும் வகையிலும் தேர்ந்துப் படிக்கும் வகையிலும் இத்தளத்தில் அமைக்கப் பெற்றுள்ளன.

இந்நூலகத்தில் இலக்கண நூல்கள், சங்க இலக்கியம், பதினெண் கீழ்க்கணக்கு, காப்பியங்கள், சமய இலக்கியங்கள், சிற்றிலக்கியங்கள், திரட்டு நூல்கள், நெறி நூல்கள், சித்தர் இலக்கியங்கள், இருபதாம்

நூற்றாண்டுத் தமிழ் இலக்கியங்கள் (கவிதை), இருபதாம் நூற்றாண்டுத் தமிழ் இலக்கியங்கள் (உரைநடை), நாட்டுப்புற இலக்கியங்கள், சிறுவர் இலக்கியங்கள் முதலானவை இடம் பெற்றுள்ளன.

இன்றைய நிலையில் தமிழகத்தில் உள்ள அனைத்துப் பல்கலைக்கழகங்களிலும் பாடமாக அமையப் பெற்றதும், பரவலாகப் பயன்பாட்டில் உள்ளதுமான அனைத்து நூல்களும் இதன் தளத்தில் உள்ளன.

சிறப்பம்சங்கள்

- இலக்கண, இலக்கிய நூல்களுக்கு ஒரே நேரத்தில் ஒன்றுக்கு மேற்பட்ட உரைகள் கிடைக்கின்றன.
- திருக்குறளுக்குத் தமிழக முதல்வர் கலைஞர் கருணாநிதி அவர்களின் உரை உட்பட எழுவரின் உரைகள் கிடைக்கின்றன.
- சங்க இலக்கிய பாடுபொருள்கள் எண், சொல், பக்கம், பாடினோர், வள்ளல்கள், மன்னர்கள், திணை, கூற்று, பாடல் முதற்குறிப்பு, மரங்கள், செடிகள், கொடிகள், தானியங்கள், பழங்கள், விலங்குகள், பறவைகள், மீன்கள் என்னும் அடிப்படையில் வேண்டிய செய்திகளை உடனடியாகத் தேடிப் பெறுகின்ற வகையில் உள்ளன.
- சொல் தேடல், எண் தேடல், பக்கம் தேடல் போன்ற தேடுதல் வசதிகள் உள்ளன.
- அகராதிகளில், தமிழ்ச் சொற்களுக்கு இணையான ஆங்கிலச் சொற்களையும், ஆங்கிலச் சொற்களுக்கு இணையான தமிழ்ச் சொற்களையும் தேடிப் பெறுகின்ற வசதி உள்ளது.
- தமிழர்களின் பண்பாட்டுக் கூறுகளை உலகத்தார்க்கு எடுத்துக்காட்டும் வகையில் அமைந்த சைவ, வைணவ, இசுலாமிய, கிறித்துவ, சமணக் கோயில்களின் ஒலி, ஒளிக் காட்சிப் பதிவுகள் மற்றும் பரத நாட்டியம், பொம்மலாட்டம், காவடியாட்டம், மயிலாட்டம், நாதஸ்வரம், ஜல்லிக்கட்டு முதலான பண்பாட்டுக் காட்சியகம் வியக்கத்தக்க வகையில் இடம் பெற்றுள்ளது.
- திருத்தலங்கள் என்னும் வாசையில் 14 சமணத் தலங்கள், 101 சைவத் தலங்கள், 93 வைணவத் தலங்கள், 9 இசுலாமியத் தலங்கள், 13 கிறித்துவத் தலங்கள் காட்சியாக்கப்பட்டுள்ளன.
- தேவாரப் பாடல்களை இசையுடன் கேட்கும் வசதி உள்ளது.
- அனைவரும் படித்து விளங்கிக் கொள்ள வசதியாக கடின நூல்கள் எளிய முறையில் பதம் பிரித்துத் தரப்பட்டுள்ளன.

இந்திய மொழிகளின் நடுவண் நிறுவனம்

மைய அரசு நிறுவனமான இது, தமிழ் நூல்களை அதன் பழைமை குன்றாமல் அதாவது மூலப் பிரதியில் உள்ளவாறே இணையவழியில் அளிப்பதற்கான முயற்சியில் முனைப்புடன் செயல்பட்டு வருகிறது. தொல்காப்பிய நூற்பாக்கள் சிலவற்றையும், சங்க இலக்கியப் பாடல்கள் சிலவற்றையும் இசையுடன் கேட்கும் வகையில் இத்தளம் அமைந்துள்ளது. பாடல்களை மட்டுமின்றி நூற்பாக்களைக்கூட இசையுடன் வழங்க முற்பட்டுள்ள இந்நிறுவனத்தின் முயற்சிப் பாராட்டுதற்குரியதாகும்.

மதுரைத் திட்டம்

உலகளாவிய தமிழர்கள் இணையம்வழி ஒன்றுகூடி தமிழ் இலக்கியங்களின் மின் பதிப்புகளை உருவாக்கி அவற்றை இணையம்வழி உலகெங்கிலும் உள்ள தமிழர்களும் தமிழார்வலர்களும் இலவசமாகப் பெற வசதி செய்யும் ஒரு தமிழ் இலக்கிய மின்பதிப்புத் திட்டமே மதுரைத் திட்டம். இதனை, சுவிட்சர்லாந்து நாட்டில் வாழும் தமிழர் கே.கல்யாணசுந்தரம் 1998ஆம் ஆண்டுத் தமிழர் திருநாளன்றுத் தொடங்கினார். இத்திட்டம் சனவரி 2008இல் தனது பத்தாம் ஆண்டு நிறைவு விழாவைக் கொண்டாடியது.

எந்த ஒரு சமூகக் கலாச்சாரத்திற்கும் அதன் இலக்கியங்களே சிறந்த ஆதாரங்களாகும். அதனை உட்யவாறு பாதுகாத்து உலகம் முழுவதும் வாழ்வோர் பகிர்ந்து கொள்ளும் வகையிலும் எதிர்காலச் சந்ததியினருக்குக் கொண்டு செல்லும் வகையிலும் தொடங்கப்பட்டதே இத்திட்டம் ஆகும். இத்தளத்தில் தமிழ் நூல்கள் ஏறக்குறைய 350 நூல்கள் இடம் பெற்றுள்ளன. தொடக்கத்தில் டிஸ்கிக் குறியீட்டில் இருந்த நூல்கள் தற்பொழுது ஒருங்குறியீட்டு முறையிலும் கிடைக்கின்றன. இதில் இடம் பெற்றுள்ள நூல்களை அகர வாசை, நூல் வாசை, கால வாசை என்னும் அடிப்படையில் காணலாம்.

இத்திட்டத்தின் சிறப்பம்சம் யாதெனில், யார் வேண்டுமானாலும் தமிழ் இலக்கியங்களை மின்பதிப்புச் செய்து இவர்களின் அனுமதியோடு அம்மின் தொகுப்பில் சேர்த்துக் கொள்ளலாம். மேலும், இத்தொகுப்பில் உள்ள நூல்கள் ஒருங்குறியீட்டு முறையிலும் கிடைப்பதனால் எழுத்துரு பிரச்சினை இல்லாமல் நூல்களைப் படிக்க முடியும். ஆனால், இத்தொகுப்பில் வேண்டிய தகவல்களைத் தேடிப் பெறும் வசதி ஒருசில நூல்களுக்கு மட்டுமே டிஸ்கிக் குறியீட்டில் உள்ளது. தமிழ் இணையப் பல்கலைக்கழக தளத்தில் வேண்டிய தகவல்களைத் தேடிப் பெறுவது போல இத்திட்டத்தில் தகவல்களைப் பெற இயலாது. ஆனால் தமிழ் இணையப் பல்கலைக்கழக தளத்தில் நூல்கள் ஒருங்குறியீட்டில் இல்லாமையால் அதனைப் படிப்பதில் எழுத்துருப் பிரச்சினை ஏற்படுகிறது.

கனடாவில் ஒரு குளிர் கால இரவில் நான் வீட்டை விட்டு ஓர் அடிகூட நகராமல் எட்டுத்தொகையில் ஏழாவதான நெடுநல்வாடையை என் கணினியில் இறக்கி நான் இலவசமாகப் படித்தேன். இது எப்படி சாத்தியமானது? இந்த தொண்டர்களின் உழைப்புக்கு விலை போட முடியுமா? என்னுடைய கணக்குப் பிரகாரம் ஒரு மில்லியன் டாலருக்கு அதிகமாவே வந்தது என்று கவிஞர் அ.முத்துலிங்கம் மதுரைத் திட்டத்தின் பணியைப் பாராட்டுகிறார். (காண்க: தொன்மையில் இல்லை, தொடர்ச்சியில், ezhilnila.com)

சென்னை நூலகம்

2006ஆம் ஆண்டு செப்டம்பர் மாதம் வணிக நோக்கில் இது தொடங்கப்பட்டது. ஆயினும் தமிழ் நூல்களை ஒருங்குறியீட்டு முறையில் இலவசமாகப் பார்வையிடும் வசதியை இத்தளம் வழங்குகிறது. இதில் பழந்தமிழ் நூல்கள் முதல் அண்மைக்கால நூல்கள் வரை இடம் பெற்றுள்ளன.

இதில் எட்டுத் தொகை, பத்துப்பாட்டு, பதினெண் கீழ்க்கணக்கு நூல்களும் சிலப்பதிகாரம், சீவகசிந்தாமணி, வளையாபதி, குண்டலகேசி ஆகிய ஐம்பெரும் காப்பிங்களும் இடம் பெற்றுள்ளன. யாப்பருங்கலக்காரிகை, நால்வர் நான்மணிமாலை, திருவிசைப்பா, திருமந்திரம், திருவாசகம், திருக்களிற்றுப்படியார், திருவுந்தியார், கந்தர் அலங்காரம், கந்தர் அந்தாதி, திருப்புகழ் முதலான நூல்களும் இத்தளத்தில் கிடைக்கின்றன. இவற்றை என் அடிப்படையில் தேடிப்பெற இயலும்.

இவையேயன்றி, கம்பர் திருஞானசம்பந்தர், திரிகூடராசப்பர், குமரகுருபரர், ஓளவையார், பாரதியார், பாரதிதாசன், பேரறிஞர் அண்ணா, மு.வரதராசன், ந.பிச்சமூர்த்தி ஆகியோரின் படைப்புகளும் இதில் கிடைக்கின்றன.

விருபா.காம்

நூல்கள், தமிழ் ஆய்வுகள் தொடர்பான பல அரிய தகவல்கள் இத்தளத்தில் கிடைக்கின்றன. இதுதான் இணையத்தில் தமிழில் வெளிவந்த முதல் தகவல் திரட்டு ஆகும். தமிழில் வெளிவந்துள்ள புத்தகங்கள், ஆய்வேடுகள், சிற்றிதழ்கள், எழுத்தாளர்கள், பதிப்பகங்கள், புத்தகப் பிரிவுகள், மதிப்புரைகள், நிலவரம் ஆகிய தலைப்புகளில் இத்தளத்தில் செய்திகள் கிடைக்கின்றன. இதனை வடிவமைத்துப் பராமாத்து வருபவர் டி.குமரேசன் என்பவர் ஆவார்.

அதிக பொருட் செலவையும் பொருட்படுத்தாமல் இவர் இத்தளத்தைத் தொடர்ந்து நடத்தி வருகிறார் என்பதும் இத்தகைய முயற்சித் தமிழில் புதிது என்பதும் ஆங்கில மொழியில் அமைந்துள்ள இணைய தளத்திற்கு இணையான வகையில் இது அமைந்துள்ளது என்பதும் மிகவும் பாராட்டத் தக்கதாகும்.

இத்தளத்தில் 30-04-2010 வரை 3027 புத்தகங்கள், 1329 எழுத்தாளர்கள், 588 பதிப்பகங்கள், 107 புத்தகப் பிடிவுகள், 206 மதிப்புரைகள், 67 சிற்றிதழ்கள் பற்றிய விபரங்கள் இடம் பெற்றுள்ளன.

தரவுகளைப் பல்வேறு தேடுதல் வசதிகளின் மூலம் பெறலாம். சான்றாக, புத்தகங்களை எழுத்தாளர், பதிப்பகம், ஆண்டு, பிடிவு ஆகிய அடிப்படையில் தேடிப் பெறலாம். ஒரு புத்தகத்தைத் தேடுவதாக வைத்துக் கொண்டால் அதன் பதிப்பு, விலை, பக்கங்கள், ISBN எண், பிரிவு, எழுத்தாளர், பதிப்பகம், முகவரி இணையதளம் உள்ளிட்ட அனைத்துத் தகவல்களும் முழுமையாக இருக்கும். இதன் மூலம் எந்தப் புத்தகம் எங்குக் கிடைக்கும் என்பதை எளிதில் தொகுந்து கொள்ளலாம். இது ஆராய்ச்சியாளர்களுக்குப் பேருதவியாக இருக்கும்.

அதுபோல் பதிப்பகங்களைத் தேடுதலாக வைத்துக் கொண்டால், அகர வாசை, பதிப்பகம் அமைந்துள்ள நாடு மற்றும் இடம் என்ற வரிசை அடிப்படையில் தேடிப் பெறலாம். தமிழில் வெளி வந்துள்ள முனைவர் பட்டம், ஆய்வியல் நிறைவுப் பட்டம் ஆகியவை நெறியாளர்கள், ஆய்வாளர்கள், பட்டம் வழங்கிய பல்கலைக்கழகம் ஆகிய தகவல்களுடன் அளிக்கப்படுகின்றன. இது தமிழாய்வுலகிற்கு மிகப் பொருடைய நன்மையாகும். குறிப்பாக ஆய்வுத் திருட்டைத் தடுக்கப் பேருதவியாக இது இருக்கும்.

இங்ஙனம் மிகச் சிறப்பான முறையில் நூல்கள் பற்றிய தகவல்களைத் திரட்டித் தருகின்ற வகையில் இத்தளம் அமைந்துள்ளதால் ஆய்வாளர்கள் நூல்கள் கிடைக்கும் இடம் பற்றிய விபரங்களை எளிதில் அறிந்து கொள்ளலாம். அது போல் நூல் பற்றிய மதிப்புரைகளும் இடம் பெற்றிருப்பதால் அந்நூல்கள் பற்றி அறிந்துகொண்டு அதன் பின்பு விரும்பினால் அவற்றை வாங்கலாம். தமிழ் நூல் பற்றிய தகவல் திரட்டுகளை இத்தளம் வழங்குவது போல வேறு எத்தனாமும் வழங்குவதாகத் தெரியவில்லை.

தற்பொழுது 1850 முதல் 1928க்கு இடைப்பட்ட காலத்தில் வெளியான புத்தகங்களில் விருபா குமரேசனாருக்குக் கிடைத்தவற்றை மின்நூல்களாக மாற்றி இதில் இணைத்து வருகிறார். இதுரை 41 புத்தகங்களை தனது தளத்தில் இணைத்துள்ளார். இவற்றை pdf கோப்புகளாக எளிதில் தரவிறக்கம் செய்து கொள்ளலாம். இங்ஙனம் இணைக்கப்பட்ட புத்தகங்கள் பற்றிய ஓர் அறிமுகத்தையும் விரைவில் இதில் இவர் வழங்க உள்ளார்.

இத்தளம் தம் பணியை மேலும் விரிவு படுத்தியும், செம்மைப்படுத்தியும் வழங்குமேயானால் தமிழ் கூறும் நல்லுலகம் இத்தனி மனிதருக்கு நன்றி சொல்ல கடமைப்பட்டதாகும். ஏனெனில், பல்கலைக்கழகங்கள் செய்ய வேண்டிய பணியை தனி மனிதர் செய்து சாதனை படைத்துள்ளார்.

நூலகம்.நெட்

ஈழத்து நூல்களையும், இதழ்களையும் மின் வடிவமாக்கி பாதுகாத்து அனைவரும் எளிதில் பார்வையிடும் வகையில் இத்தளம் செயல்பட்டு வருகிறது. இது ஓர் இலாப நோக்கமற்ற தன்னார்வக் கூட்டு முயற்சியாகும். இதன் பணிகள் ஏறக்குறைய மதுரைத்திட்டம் போலவே அமைந்துள்ளது.

இந்நூலகம் தொடக்கத்தில் ஈழ நூல் என்ற பெயரில் இருந்தது. தி.கோபிநாத், மு.மயூரன் ஆகியோர் 2005 சனவரியில் இதனைத் தொடங்கினர். சில காரணங்களால் நின்று போன இத்திட்டம் 2006இல் தமிழர் திருநாளன்று 100 மின் நூல்களுடன் மீண்டும் உதயமானது. வாரம் ஒரு மின் நூல் என்னும் குறிக்கோளுடன் செயற்பட்டு வரும் இத்தளத்தில் 30-04-2010 வரை 6072 நூல்கள் இடம் பெற்றுள்ளன.

இதில் ஆவண வகைகள், பகுப்புகள், பட்டியல்கள் என மூன்று பெரும் தலைப்புகள் இடம் பெற்றுள்ளன. இதில் ஆவண வகைகள் என்பதில், நூல்கள், சஞ்சிகைகள், பத்திரிகைகள், பிரசுரங்கள், ஆய்வேடுகள் என்னும் பகுப்புகள் உள்ளன. பகுப்புகள் என்பதில் எழுத்தாளர்கள், வெளியீட்டு ஆண்டு, பதிப்பகங்கள், நூல் வகை ஆகியவை இடம் பெற்றுள்ளன. பட்டியல்கள் என்பதில், நூறு எண்கள் கொண்ட தொகுப்புகளாக 6072 நூல்கள் உள்ளன. இத்தலைப்புகளில் மட்டுமின்றி பிற மின் நூல்கள் என்ற தேடலில் பிற நிறுவனங்களின் மின் நூல்களையும் இதில் காணலாம்.

ரோஜா முத்தையா ஆராய்ச்சி நூலகம்

ரோஜா முத்தையா செட்டியாரின் நினைவால் சிகாகோ பல்கலைக்கழகம் இந்நூலகத்தை 1994 முதல் நடத்தி வருகிறது. ஏறக்குறைய ஒரு இலட்சத்திற்கும் மேற்பட்ட அரிய நூல்களும் இதழ்களும் இந்நூலகத்தில் உள்ளன. நூலாசிரியர், நூலின் தலைப்பு, நூல் வெளிவந்த ஆண்டு என்னும் அடிப்படையில் இந்நூலகத்தில் உள்ள பட்டியல்களை ஆராய்ச்சியாளர்கள் இணையத்தில் தேடிப்பெறும் வகையில் 1996 முதல் செயற்படுத்தி வருகிறது. ஆய்வாளர்கள் இருந்த இடத்தில் இருந்து கொண்டே ஆய்வுக்குத் தேவையான நூல்கள் இந்நூலகத்தில் உள்ளனவா என்பதை அறிந்து கொண்டு வேண்டிய நூல்கள் இருந்தால் இந்நூலகத்திற்குச் சென்று பயனடையலாம்.

விக்கிபீடியா

தமிழில் கலைக்களஞ்சியம் இருப்பது போல இணையக் களஞ்சியமாக வீறுநடை போட்டு வருவது விக்கிபீடியா ஆகும். அமெரிக்காவை மையமாகக் கொண்டு செயல்பட்டு வரும் இது ஓர் இலாப நோக்கமற்ற அமைப்பாகும். இது தமிழ் உள்ளிட்ட 54 உலக மொழிகளில் தகவல்களை வெளியிட்டு வருகிறது. தமிழில் 30-04-2010 வரை 22324 கட்டுரைகள் இடம் பெற்றுள்ளன. இதன் விக்கிசனாடி தமிழ் அகர முதலி போன்றது ஆகும். கட்டற்ற அகர முதலியான இதில் இதுகாறும் 1,13,102 தமிழ்ச் சொற்கள் இடம் பெற்றுள்ளன. இது உலகத் தமிழ்ச் செம்மொழி மாநாட்டினையொட்டித் தமிழக அரசுடன் இணைந்து விக்கிபீடியா தகவல் பக்கங்கள் போட்டி ஒன்றினையும் ஏற்பாடு செய்துள்ளது. இதன் மூலம் தனது தகவல் பக்கங்களை வெகுவாக இது உயர்த்திக் கொள்ளும் என்று எதிர்பார்க்கப்படுகிறது. இத்தளத்தில் தமிழ் தொடர்பான அனைத்துத் தகவல்களையும் மிக விரைவில் காண முடியும் என்னும் நம்பிக்கைப் பலருக்கும் உண்டு.

முடிப்புரை

முன்பெல்லாம் தமிழறிஞர்களும், தமிழ் ஆர்வலர்களும், ஆய்வாளர்களும் தங்களுக்கு வேண்டிய நூல்களைத் தேடி ஊர் ஊராக அலைந்து அல்லல் பட்டனர். ஆனால் இன்று அத்துன்பங்கள் இணையம் வழியாகத் தவிர்க்கப்பட்டுள்ளன. வீட்டில் கணினியின் முன் அமர்ந்து அனைத்துத் தமிழ் நூல்களையும் மேலே காட்டியது போன்ற இணைய தளங்களில் காணும் பேறு அனைவருக்கும் வாய்க்கப் பெற்றுள்ளது. அப்பேற்றைக் கண்டறிவதும், பயன் பெறுவதும் ஒவ்வொருவரின் தலையாயக் கடமையாகும். இதனை உணர்ந்தே உலகத் தமிழ்ச் செம்மொழி மாநாடும், அதன் பகுதியாக ஒன்பதாவது உலகத் தமிழ் இணைய மாநாடும் சீரிய முறையில் நடந்தேறி வருவது பாராட்டுதலுக்குரியதாகும்.

பயன் கொள்ள வேண்டிய இணைய தளங்கள்

1. www.tamilvu.org
2. www.ciil.org
3. www.tamil.net/projectmadurai
4. www.chennaiLibrary.com
5. www.viruba.com
6. www.noolaham.net
7. www.lib.uchicago.edu/e/su/southasia/rmrl.html
8. www.ta.wikipedia.org
9. www.ezhilnila.com

மின்வழி அருங்காட்சியகம்

(Electronic Museum)

மறைமலை இலக்குவனார்

கட்டுரையின் நோக்கம்

மின்வழி அருங்காட்சியகம் அல்லது நிகர்நிலை அருங்காட்சியகம் அல்லது எண்ம அருங்காட்சியகம் என்பது இணையதளத்தில் உருவாக்கப்படும் இணையவழி அருங்காட்சியகத்தைக் குறிக்கும் சொல் ஆகும். இத்தகைய அருங்காட்சியகம், நடைமுறை உலகில் நிறுவப்பட்டுள்ள அருங்காட்சியகங்களின் செய்திகள், காட்சிகள், ஒலிகள் அனைத்தையும் தொகுத்து வழங்கும் அருங்காட்சியக மாதிரிகளாகவும் இருக்கலாம்; அல்லது நிகர்நிலை உலகில் (இணையதளத்தில்) உருவாக்கப்பட்டுத் தகவல் களஞ்சியங்களாகவும் வலைக்காட்சிவிருந்தாகவும் விளங்கும் இணைய அருங்காட்சியகங்களாகவும் அமையலாம். இவ்விரு பிரிவிலும் விளங்கும் மின்வழி அருங்காட்சியகங்கள் குறித்த செய்திகளை வழங்குவதும், தமிழ்க்கணினிப் பயன்பாட்டை மேம்படுத்தும் வகையில் இத்தகைய அருங்காட்சியகங்களை அமைக்கச் சில பரிந்துரைகளை வழங்குவதும் இக் கட்டுரையின் நோக்கங்களாகும்.

மின்வழி அல்லது இணையவழி அருங்காட்சியகங்கள் - வகைப்பாடு

அருங்காட்சியகங்களுக்கு வரும் பார்வையாளர்கள் அவற்றில் அமைந்துள்ள காட்சிக்கூடங்கள் பற்றியோ அவற்றுள் காத்து வைக்கப்பட்டுள்ள படிமங்கள், அகழ்வாய்வின் மூலம் கண்டெடுக்கப் பட்டுள்ள தொல்லியல் சான்றுகள் பற்றியோ விரிவுறத் தெரிந்திருப்பார்கள் எனக் கூறவியலாது. இவை குறித்த விளக்கங்கள் அடங்கிய கையேடுகள் அருங்காட்சியகங்களில் வழங்கப்படுவதால், பார்வையாளர்கள் அவற்றைப் படித்து அருங்காட்சியகத்தை முழுதுமாகப் பயன்கொள்ள இயல்கிறது. இதனைவிட இன்னும் விரைவாகவும் விரிவாகவும் அருங்காட்சியகத்தையும் அதன் காட்சிப்பொருள்களையும் தெரிந்துகொள்ளத் துணைபுரியும் வகையில் இணையவழி அருங்காட்சியகங்கள் பயன்படுகின்றன. இவை மிகவும் விரிந்து பரந்து பல்வேறுவகைகளில் அமைகின்றன.

அ) ஏற்கெனவே அமைந்துள்ள அருங்காட்சியகத்தின் வலையீடுகள்

ஆ) நடைமுறை உலக அருங்காட்சியகத்தின் நிகர்நிலைச் சுற்றுலா

இ) நிகர்நிலை உலக அருங்காட்சியகத்தின் நிகர்நிலைச் சுற்றுலா

ஈ) பயனாளர்களின் துணையுடன் அமைக்கப்படும் நிகர்நிலை

அருங்காட்சியகங்கள்

உ) அருங்காட்சியக வலைத்தளங்களில் விளங்கும் கல்விசார்

விளையாட்டுக் குறித்தவை

ஊ) எண்ணற்ற கண்காட்சிகளை ஒருங்கிணைக்கும் வலைத்தளங்கள்

எ) தொழில்நுணுக்கத் துணையுடன் அருங்காட்சியகத்தைப் பயன்கொள்ள

வழிவகுப்பன.

இவைகுறித்து ஒரு சில செய்திகள் இங்குக் குறிக்கத் தக்கன.

அ) ஏற்கெனவே அமைந்துள்ள அருங்காட்சியகத்தின் வலையீடுகள்

இதற்கு எடுத்துக்காட்டாகச் சென்னையில் அமைந்துள்ள அரசு அருங்காட்சியகத்தின் வலையீட்டைக் கூறலாம். இதன் வலைத்தள முகவரி:

<http://www.chennaiuseum.org/draft/index.htm>

சென்னை அரசு அருங்காட்சியகத்தின் தளவிளக்க வரைபடம், அருங்காட்சியகச் செய்தி, காணொளிநறுக்குகள், அருங்காட்சியகவரலாறு, பொதுத்தகவல்கள், காட்சிக் கூடங்கள், பல்வேறுதுறைகள்/பிரிவுகள், வெளியீடுகள், கல்விசார் நிகழ்ச்சிகள், மாவட்ட அருங்காட்சியகங்கள், எதிருட்டு ஆகியவற்றுக்கான இணைப்புகளும், நிகர்நிலைச் சுற்றுலாவுக்கான இணைப்பும் கொண்டு இதன் முகப்புப்பக்கம் அமைந்துள்ளது. தொல்லியல், கலை, மானிடவியல், நாணயவியல், விலங்கியல், தாவரவியல், நிலப்பொதியியல், சிறுவர் அருங்காட்சியகம், வேதிமப் பாதுகாப்பு ஆகியவற்றுக்கான பிரிவுகளுள் ஒன்றினைச் சொடுக்கி நுழைந்தால் நிறையத் தகவல்கள், பட விளக்கங்கள் ஆகியவற்றைப் பெறலாம். 25/04/2010 வரை இவ் வலைத்தளத்தைப் பார்வையிட்டோர் 2,84,611 பேர் என்னும் குறிப்பு மன நிறைவளிக்கிறது.

ஆ) நடைமுறை உலக அருங்காட்சியகத்தின் நிகர்நிலைச் சுற்றுலா

சென்னை அரசு அருங்காட்சியகத்தின் முகப்புப்பக்கத்திலேயே நிகர்நிலைச் சுற்றுலாவுக்கான இணைப்பு வழங்கப்பட்டுள்ளது. எனினும், நிகர்நிலைச் சுற்றுலாவுக்கு அடித்தளமாக அமைவது *விஆர்எம்எல்* என்றழைக்கப்படும் *நிகர்நிலை உருவமை மொழி* ஆகும். முப்பரிமாண மாதிரிகளையும் செயலியக்கப் படங்களையும் இதன் துணைக்கொண்டுதான் உருவாக்குகிறார்கள். *காசுமாபினேயர்* எனப்படும் மென்பொருளை நம் கணினியில் பதிவு செய்துகொண்டால்தான் *விஆர்எம்எல்* துணைக்கொண்டு அமைக்கப்பட்ட நிகர்நிலைச்சுற்றுலாவைக் கண்டு பயன்பெறமுடியும். *காசுமாபினேயரைப்* பதிந்துகொள்வதும் எளிமையாக இல்லை. பாரிசு நகரத்தில் அமைந்துள்ள லூவர் அருங்காட்சியகத்திற்குரிய நிகர்நிலைச்சுற்றுலா மிகவும் சிறப்பாக அமைக்கப்பட்டுள்ளது. இதனைக் காண்பதற்கு *மேக்ரோமீடியா பிளாசுபினேயர்* இருந்தால் போதுமானது.

<http://www.louvre.fr/llv/commun/home.jsp>

என்னும் வலைமுகவரிக்குச் சென்றால் மிக எளிமையாக ஒரு பயன்மிக்க சுற்றுலாவுக்குச் சென்றுவந்த நிறைவு பெறலாம். பிரெஞ்சு, ஆங்கிலம், சப்பானிய மொழிகளில் அமைந்துள்ள விளக்கங்களும், முப்பரிமாணப் படங்களும், செயலியக்கப் படங்களும் ஒரு புதியஉலகை நம் கண்முன் நிறுத்துகின்றன என்றும், புதிய கல்வியை வழங்குகின்றன என்றும் கூறலாம்.

இ) நிகர்நிலை உலக அருங்காட்சியகத்தில் நிகர்நிலைச் சுற்றுலா

உருகுவே நாட்டுக் கலைகளைப் பற்றிய நிகர்நிலை அருங்காட்சியகம் முழுமையும் இணையதளத்திலேயே அமைக்கப்பட்டதாகும். எனினும் கட்டடம், அறைகள், பல்வேறு தளங்கள் படமாகவே அமைக்கப்பட்டுப் பல்வேறு காட்சிக்கூடங்களைக் கொண்டமைக்கப்பட்டுள்ளது. பின்வரும் வலைமுகவரிக்குச்

சென்றால் அந்த அருங்காட்சியகத்தைக் காணலாம்.

<http://muva.elpais.com.uy>

கீழ்த்தளத்தில் தொடங்கிப் பல்வேறு தளங்களுக்கும் நாம் செல்வதைப் போன்ற ஒரு தோற்றத்தை மிக நேர்த்தியாக அமைத்து, ஒரு பெரிய மண்டபத்தின் முன் நாம் அமர்ந்திருப்பதைப் போலவும், பல்வேறு இடங்களுக்குச் சென்று பல ஓவியங்களைப் பார்ப்பதைப் போலவும் வலைப்பக்கங்களை அமைத்துள்ள வனப்பு போற்றத்தக்கது.

ஈ) பயனாளர்களின் துணையுடன் அமைக்கப்படும் நிகர்நிலை அருங்காட்சியகங்கள்

உலகில் நிறுவப்பட்டுள்ள அருங்காட்சியகங்களும் இணையவெளியில் மட்டுமே காணக்கூடிய அருங்காட்சியகங்களும் தமது முகப்புப்பக்கங்களில் பார்வையாளர்களுக்கு வேண்டுகோள் விடுத்து அல்லது

குறிப்பிட்ட தன்மையிலான கலைப் படைப்புகளை வரவேற்றுப் புதிய பன்னூற்றுக்கணக்கான படைப்புகளை வலைத்தளத்தில் குவிப்பதன்மூலம் கலைவிருந்து படைக்கின்றன.

<http://virtual-museum-india.blogspot.com>

என்னும் வலைப்பூ இவ்வகையில் ஒரு குறிப்பிடத்தக்க முயற்சியை மேற்கொண்டுவருகிறது. ஓவியங்களுடன் குறும்படங்களும் இந்த நிகர்நிலை அருங்காட்சியகத்தில் இடம்பெற்றுள்ளன. வரலாற்றுக்கு முற்பட்ட கலை குறித்த நிகர்நிலை அருங்காட்சியகம் முற்றிலும் இணையவெளியில் மட்டுமே அமைக்கப்பட்டுள்ள ஒன்றாகும்.

<http://vm.kemsu.ru/en/index.html>

என்னும் வலைமுகவரிக்குச் சென்றால் அரிய கலைக்காட்சிகள் பல கண்ணுக்கும் அறிவுக்கும் விருந்தாக அளிக்கப்பட்டுள்ளன. *கெமொரோவொ* அரசுப் பல்கலைக்கழகத்தின் தொல்லியல் துறையினால் உருவாக்கப்பட்டுள்ள இந்நிகர்நிலை அருங்காட்சியகம் பல்வேறு தொல்லியல் ஆய்வாளர்களின் கண்டுபிடிப்புகளைத் தொகுத்து அமைக்கப்பட்டுள்ளது. ஆங்கிலத்திலும் உருசிய மொழியிலும் செய்திகள் வழங்கப்பட்டுள்ளன.

இளைஞர்களால் உருவாக்கப்பட்டு இளைஞர்களின் கலைப்படைப்புகளை அரங்கேற்றப் பாடுபடும் நிகர்நிலை அருங்காட்சியகம் ஒன்றினைப் பின்வரும் வலை முகவரியில் காணலாம்.

<http://ungeslaboratorierforkunst.dk/index.asp?key=1>

உருவாக்கு, தேடிக்கண்டுபிடி, பங்குபெறு என்னும் மூன்று முழக்கங்களுடன் இதன் முகப்புப்பக்கம் ஆங்கிலத்திலும் டேனிசு மொழியிலும் அமைந்துள்ளது. உருவாக்கு என்னும் இணைப்பைச் சொடுக்கித் தன்விவரக்குறிப்பை அளித்தால் பின்னர் படைப்புகளை இந்த அருங்காட்சியகத்திற்கு வழங்கலாம்.

மியூசியோ தே பெசோவா என்னும் பெயரில் அமைந்துள்ள போர்த்துகீசிய நிகர்நிலை அருங்காட்சியகம் போர்த்துகீசிய மொழியிலும் ஆங்கிலத்திலும் இயங்குகிறது. 'தனி மனிதனின் அருங்காட்சியகம்' என்னும் தலைப்பிலமைந்த இந்த நிகர்நிலை அருங்காட்சியகம் 'அனைத்து மனிதர்களின் வரலாறும் மனித இன வரலாற்றின் பகுதியே' என்னும் கருத்துடன் ஒவ்வொருவரின் வாழ்க்கையையும் -நிகழ்ச்சிகள், குறிப்புகள், படங்கள்- பதிவுசெய்துகொள்ள உதவுகிறது. ஆங்கிலத்திலமைந்துள்ள பகுதியைக் காட்டிலும் போர்த்துகீசிய மொழியிலமைந்துள்ள பகுதி விரிவாகவும் பல்வேறு மனிதர்களின் வாழ்க்கைவரலாற்றுத் தொகுப்பாகவும் அமைந்துள்ளது. இக்காலச் சமுதாயத்தின் பன்முகப்பாங்கான வரலாற்றைச் சுவையுடனும் மிகையின்றியும் வழங்கும் சமூக ஆவணமாக இந்த நிகர்நிலை அருங்காட்சியகம் விளங்குகிறது.

உ) அருங்காட்சியக வலைத்தளங்களில் விளங்கும் கல்விசார் விளையாட்டுக் குறித்தவை

பிரிட்டிசு அருங்காட்சியகத்தின் பண்டைய கிரேக்கம் குறித்த வலைத்தளத்தில் காட்டப்பட்டுள்ள ஒவ்வொரு தலைப்பிலும் ஒவ்வொரு விளையாட்டு அமைந்துள்ளது. கோயில்களைக் கட்டுவது, உடைந்த கப்பல்களைக் கண்டறிவது, கடலுக்கடியில் அமைந்துள்ள கருவூலங்களைத் தேடுவது போன்ற விளையாட்டுகள் மனமகிழ்வுக்கும் புதிய செய்திகளை அறிந்துகொள்வதற்கும் பயன்படுகின்றன. பார்வைக்கு:

<http://www.ancientgreece.co.uk/menu.html>

கனடா நிகர்நிலை அருங்காட்சியகத்திலும் இத்தகைய கல்விசார் விளையாட்டுகள் இலவசமாக வழங்கப்பட்டுள்ளன. இவ்வலைத்தளம் குறித்து அடுத்த பிரிவில் காணலாம்.

ஊ) எண்ணற்ற கண்காட்சிகளை ஒருங்கிணைக்கும் வலைத்தளங்கள்

சிங்கப்பூர் அரசு தேசிய மரபு வாரியம் என்னும் அமைப்பின்வழி அருங்காட்சியகத்தை விளக்கும் நிகர்நிலை அருங்காட்சியகத்தை அமைத்துள்ளது. 2007-ஆம் ஆண்டு 'இணையவழிச் சிங்கப்பூர் கலைத்தொகுப்புகள்'

என்னும் தலைப்பில் இணையதளத்தில் அந்த வாரியம் தொகுத்த கலைப்படைப்புகள், கலைப்பொருள்கள் ஆகியவற்றை முப்பரிமாணப் படங்களுடன் வழங்கியது. 2010-இல் இத்தொகுப்பு ஆறுமடங்காகப் பெருகியுள்ளதாக அண்மைக் குறிப்பு ஒன்று தெரிவிக்கிறது. தேசிய மரபு வாரியம் இணையத்தின்மூலம் முப்பத்தெட்டாயிரம்(38,000) கலைப் பொருள்களை மக்கள் எந்நேரமும் பார்வையிட வகை செய்துள்ளது. அருங்காட்சியகத்திலுள்ள கலைப்பொருள்களை மட்டுமின்றிக் கடந்த ஆண்டுகளில் நடைபெற்ற கண்காட்சிகளையும் மக்கள் இணையத்தின் மூலம் பார்வையிட இந்த நிகர்நிலை அருங்காட்சியகம் துணைபுரிகிறது.

<http://www.nhb.gov.sg/WWW>

என்னும் தளத்திற்குச் சென்றால் சிங்கப்பூர்த் தேசிய மரபு வாரியம் இயக்கும் எட்டு அருங்காட்சியகங்களையும் பார்வையிடலாம். இவை தவிரப் பல்வேறு கண்காட்சிகள் பற்றியும் தெரிந்துகொள்ளலாம்.

மிகப்பெரிய நிகர்நிலை அருங்காட்சியகமாகக் கருதப்படுவது கனடா நிகர்நிலை அருங்காட்சியகம்தான். இது கனடா நாட்டின் மூவாயிரத்து ஐம்பத்துநான்குக்கு மேற்பட்ட அருங்காட்சியகங்களை ஒருங்கிணைத்து நம் பார்வைக்கு வழங்குகிறது; இலவச இணையதள விளையாட்டுகள் பலவற்றிற்குத் தளமாகவும் அமைந்துள்ளது; கல்விசார் தகவல் களஞ்சியமாகத் திகழ்கிறது; ஐந்துஇலட்சத்து எண்பதாயிரத்துக்கும் மேற்பட்ட படிமங்களை வழங்குகிறது; அனைத்துச் செய்திகளையும் பிரெஞ்சு மொழியிலும் ஆங்கிலமொழியிலும் படிக்கும் வாய்ப்பை நல்குகிறது. <http://www.museevirtuel-virtualmuseum.ca/index-eng.jsp>

என்னும் தளம் நமக்கு மலைப்பையூட்டுகிறது.

எ) தொழில்நுணுக்கத் துணையுடன் அருங்காட்சியகத்தைப் பயன்கொள்ள வழிவகுப்பன

அருங்காட்சியகத்திற்கு வரும் பார்வையாளர்களுக்குக் கலைப்பொருள்கள் பற்றியும், தொல்லியல் கண்டுபிடிப்புகள் பற்றியும் விளக்கம் வழங்க வழிகாட்டி உடனவருவது மிகப் பழைய நிகழ்முறையாகும். ஒவ்வொருவரும் பயன்படுத்தும் வண்ணம் தனியானகை எண்மவழிகாட்டி கண்டுபிடிக்கப்பட்டுப் பார்வையாளர்களுக்கு வழங்கப்பட்டு, அவற்றின் துணையுடன் ஒவ்வொரு காட்சிமாடத்தையும் விளக்கும்முறை பின்பற்றப்பட்டது. அதன்பின்னர் மின்னியல் பல்லாடக வழிகாட்டிகள் உருவாக்கப்பட்டு, கலைக்கூடங்களைப் பற்றிய பதிவுகளை முன்கூட்டியே பதிவு செய்துவிட்டு அவற்றின் துணையுடன் விளக்கம்பெறும் முறை அறிமுகப்படுத்தப்பட்டது. இதன்பின்னர் ஒவ்வொரு காட்சிக்கூடத்திற்கும் ஓர் எண்ணிடப்பட்டது. கைப்பேசிகள் வழங்கப்பட்டு அவற்றில் ஒவ்வொரு எண்ணுக்கும் ஒவ்வொரு விளக்கவுரை பதிவுசெய்யப்பட்டு, குறிப்பிட்ட எண்ணுடைய காட்சிக்கூடத்திற்கு வந்ததும் அதற்குரிய எண்ணை அழுத்தினால் அதற்குரிய விளக்கவுரை கேட்டுக்கொள்ளும் வழி பின்பற்றப்பட்டது.

இப்போது தென்மார்க்கு கலைக்கூடத்தில் பார்வையாளர்களுக்கு ஐப்பாடு வழங்கப்படுகிறது. அருங்காட்சியகத்தில் அமைந்துள்ள காட்சிக்கூடங்கள் பற்றியும் கலைப்பொருள்கள் பற்றியும் ஒலி-ஒளி விளக்க வழிகாட்டிகள் அக் கருவிகள் மூலம் அளிக்கப்படுகின்றன. இணையதளத்தில் அருங்காட்சியகங்கள் பற்றிய விளக்கமளிக்கும் முறை இப்போது உலகெங்கும் நடைமுறைக்கு வந்துள்ளதால் அவரவரும் தமது கைக்கணினி மூலம் விளக்கமறிந்துகொள்ளும் காலம் நெருங்கிவிட்டது. அதுவரை அருங்காட்சியகத்திற்கு வரும் முன்னர் தமது வீட்டில் கணினியின் துணைக்கொண்டு தக்க அறிமுகம் பெற்றுவரலாம்.

தமிழ்க்கணினிப் பயன்பாட்டை மேம்படுத்தும் வழிமுறைகள்

நிகர்நிலை அருங்காட்சியகங்கள் இன்று உலகெங்கும் தகவல்பரிமாற்றத்திற்கும் கல்வி வளர்ச்சிக்கும் பெருந்துணை புரிகின்றன. அஞ்சல்முறை வளர்ச்சி, கணினித்துறை வளர்ச்சி, மின்னியல்துறை வளர்ச்சி ஆகியவை குறித்த நிகர்நிலை அருங்காட்சியகங்களும் நிறைய உருவாகியுள்ளன. இதுபோன்றே தொல்லியல் துறை, நாணயவியல் துறை, பண்பாடு, கலைகள் குறித்த பல்வேறு அருங்காட்சியகங்களும் இணையவெளியில் மண்டிக்கிடக்கின்றன. இச்சூழலில் தமிழ்ப்பண்பாடு, கவின்கலைகள், நுண்கலைகள், நாட்டுப்புறக்கலைகள், நாட்டுப் புறநம்பிக்கைகள், வரலாறு, எனப் பல்வேறு துறைகளில் தகவல்கள் திரட்டப்பட்டுப் பல அரிய நூல்கள் வெளிவந்துள்ளன. அவற்றின் துணையுடனும் களப்பணி ஆற்றியும் செறிந்த விவரங்களுடன் தெளிவான முறையில் கருத்துப் பரப்பும் நோக்குடன் நிறைய நிகர்நிலை அருங்காட்சியகங்களை உருவாக்கலாம்.

புலவர் வரலாறு, தலைவர் வரலாறு எனப் பொதுநிலைவாய்ந்தவையாகவோ, புகழ்பெற்ற கல்வெட்டுகள், குறிப்பிடத்தக்க வரலாற்று நிகழ்வுகள் எனக் கரு ஒன்றினை அடிப்படையாகக் கொண்டவையாகவோ இவை அமையலாம். தமிழ்ப்பல்கலைக்கழகத்தின் மேற்பார்வையில் செம்மொழி மத்திய நிறுவனத்தின் துணையுடன் தொல்லியல் சிறப்பு வாய்ந்த இடங்கள் குறித்த தரவுகள், தொன்மை வாய்ந்த கல்வெட்டுகள், பண்டை வேந்தர், புலவர் பற்றிய வரலாற்றுச் சான்றாதாரங்கள் ஆகியவற்றை ஒருங்கிணைத்து நிகர்நிலை அருங்காட்சியகங்களை உருவாக்கலாம். செய்திமூலங்களைத் திரட்டக்கூடிய வினைத்திட்டமும், தக்க அறிஞர் துணையும் வாய்க்கப்பெற்றால் ஓரளவு கணினிப்பயிற்சி உடையவர்கள் இவ்வாறு நிகர்நிலை அருங்காட்சியகங்களை உருவாக்க முடியும். ஆண்டுதோறும் உருவாக்கப்படும் நிகர்நிலை அருங்காட்சியகங்களுக்கு நிதி-நல்கை (மானியத் தொகை) வழங்குவதன்மூலமும், சிறந்தவற்றிற்குப் பரிசுகள் வழங்கிப் பாராட்டுவதன் மூலமும், இவை பெருகுதற்கும் இவற்றின் மூலம் தமிழ் மறுமலர்ச்சி தழைத்தற்கும் மத்திய மாநில அரசுகளும் அரசுசார் நிறுவனங்களும் முன்வரவேண்டும் என்பது இக்கட்டுரையாளரின் வேண்டுகோள்.

சொற்கோவை

மின்வழி அருங்காட்சியகம் - **Electronic Museum**
இணையவழி அருங்காட்சியகம் - **Online Museum**
நிகர்நிலை அருங்காட்சியகம் - **Virtual Museum**
எண்ம அருங்காட்சியகம் - **Digital Museum**
வலையீடு - **Web Presentation**
தளவிளக்க வரைபடம் - **Site Map**
எதிர்நீட்சு - **Feedback**
நிகர்நிலைச் சுற்றுலா - **Virtual Tour**
நிலப்பொதியியல் - **Geology**
வேதிமப் பாதுகாப்பு - **Chemical Preservation**
விஆர்எம்எல் - **VRML**
நிகர்நிலை உருவமை மொழி - **Virtual Reality Modeling Language**
முப்பரிமாணப் படங்கள் - **3D Images**
செயலியக்கப் படங்கள் - **Animated Images**
மின்னியல் பல்லாடக வழிகாட்டி - **Electronic multimedia guide**
ஐபாடு - **Ipod**

தமிழ் விக்கிப்பீடியா எனும் தமிழ்க் கலைக்களஞ்சியம்

தேனி.எம்.சுப்பிரமணி

ஆசிரியர், முத்துக்கமலம் இணைய இதழ் (<http://www.muthukamalam.com>)
19/1, சுசுதேவ் தெரு, பழனிசெட்டிபட்டி, தேனி - 625 531, தமிழ்நாடு, இந்தியா.
மின்னஞ்சல்: msmuthukamalam@gmail.com

முன்னுரை

உலகில் மனித அறிவுக்குத் தெரிந்த அனைத்துத் தகவல்களையும் ஒன்றாகத் தொகுக்க வேண்டும் என்கிற எண்ணம் ஏற்பட்டதன் முதிர்ச்சியாக “என்சைக்ளோபீடியா பிரிட்டானிகா” எனும் மிகப்பெரும் கலைக்களஞ்சியம் தோன்றியது. இதில் ஏற்கனவே தொகுக்கப்பட்ட தகவல்களுடன், உலகில் அவ்வப்போது ஏற்பட்டு வரும் புதிய மாற்றங்கள் குறித்த தகவல்களும் சேர்க்கப்பட்டு வருகின்றன. உலகிலுள்ள ஒவ்வொரு நாட்டிலும், அந்நாட்டின் பயன்பாட்டிலுள்ள மொழிகளில் இந்த “என்சைக்ளோபீடியா பிரிட்டானிகா”வை அடிப்படையாகவும், வழிகாட்டியாகவும் கொண்டு கலைக்களஞ்சியங்கள் உருவாக்கப்பட்டு இருக்கின்றன. இந்தக் கலைக்களஞ்சியங்களை நவீன ஊடகமான இணையத்தில் கொண்டு வரும் முயற்சியில் உருவாக்கப்பட்டதுதான் விக்கிப்பீடியா. தமிழ் உட்பட 267 மொழிகளில் உருவாக்கப்பட்டிருக்கும் விக்கிப்பீடியாவில் ஒவ்வொரு மொழியிலும் தன்னார்வப் பயனர்களின் பங்களிப்போடும் கூட்டு முயற்சியோடும் பல அரிய தகவல்களைக் கொண்ட கட்டுரைகள் தொடர்ந்து பதிவு செய்யப்பட்டு வருகின்றன. இதில் தமிழ் மொழியிலான விக்கிப்பீடியாவும் ஒரு அரிய தமிழ்க் கலைக்களஞ்சியமாக உருவாக்கப்பட்டு வருகிறது.

விக்கிப்பீடியாவின் தொடக்கம்

2001 ஆம் ஆண்டில் இணையதளம் அமைப்பதில் வல்லவரான ஜிம்மி வேல்ஸ் மற்றும் தத்துவ ஆசிரியரான லாரி சாஞ்சர் ஆகியோர் அடங்கிய குழுவினர் விக்கிப்பீடியாவைத் தொடங்கினர். ஹவாய் (Hawaii) மொழியில் விக்கி (Wiki) என்ற சொல்லுக்கு விரைவு என்று பொருள். அறிவு சார்ந்த தகவல்களை விரைவாகப் பயன்பாட்டாளர்களுக்குத் தருவதால் இணைய வழி என்சைக்ளோபீடியாவிற்கு விக்கிப்பீடியா எனப் பெயரிட்டதாக இவர்கள் தெரிவிக்கின்றனர். இந்த விக்கிப்பீடியாவிற்காக 2001 ஆம் ஆண்டில் ஜனவரி 12 ஆம் தேதியில் www.wikipedia.com என்கிற இணைய முகவரியும், ஜனவரி 13 ஆம் தேதியில் www.wikipedia.org என்கிற இணைய முகவரியும் பதிவு செய்யப்பட்டு, ஜனவரி 15 ஆம் தேதியில் “விக்கிப்பீடியா” (Wikipedia) தொடங்கப்பட்டது.

தமிழ் விக்கிப்பீடியா

விக்கிப்பீடியா அமைப்பு, 2001 ஆம் ஆண்டு மார்ச் மாதத்தில் பிரெஞ்சு மொழியிலும், மே மாதத்தில் ஜெர்மன் மொழியிலும் விக்கிப்பீடியாக்களை உருவாக்கியது. இத்துடன் விக்கிப்பீடியா தளத்தில் பிற மொழிகளில் ஆர்வமுடையவர்கள் அம்மொழிகளில் விக்கிப்பீடியாக்களை உருவாக்கிக் கொள்ளும் வசதி செய்யப்பட்டது. இவ்வசதிகளைப் பயன்படுத்தித் தங்கள் மொழிகளில் ஆர்வமுடைய பலர், அவரவர் மொழிகளுக்கான விக்கிப்பீடியாக்களைத் தொடங்கினர். இப்படி உலகின் பல்வேறு நாடுகளில் பயன்பாட்டில் இருந்து வரும் சுமார் 267 மொழிகளில் விக்கிப்பீடியாக்கள் உருவாக்கப்பட்டுள்ளன.

2003 ஆம் ஆண்டில் செப்டம்பர் 30 ஆம் தேதியில் “மனித மேம்பாடு” என்கிற தலைப்பில் ஒரு சிறு தகவல் தமிழில் இடம் பெற்றது. ஆனால் பின்னர் இது தமிழ் விக்கிப்பீடியாவின் நடைமுறைகளுக்கு ஏற்றதாக இல்லை என நீக்கப்பட்டது. இலங்கையில் பிறந்து வளைகுடா நாடான அபுதாபியில் கட்டிடப் பொறியாளராக இருந்து வரும் மயூரநாதன் என்பவர் 2003 ஆம் ஆண்டில் நவம்பர் மாதத்தில் தமிழ் விக்கிப்பீடியாவிற்கான முகப்புப் பக்கத்தை (முதற்பக்கம்) முதன்முதலாகத் தமிழில் உருவாக்கினார். அதன் பிறகு இவர் தொடர்ந்து பல கட்டுரைகளின் வாயிலாகப் பங்களிப்பு செய்து தமிழ் விக்கிப்பீடியாவின் பக்கம் உலகத் தமிழர்கள் பலரது கவனத்தையும் கொண்டு வந்தார்.

தன்னார்வப் பயனர்கள்

வணிக நோக்கமற்ற விக்கிமீடியா பவுண்டேசன் எனும் அமைப்பால் இயக்கப்படுகிறது. விக்கிப்பீடியாவில் கட்டுரைகள் மற்றும் பக்க வடிவமைப்புகளில் பங்களிப்பவர்கள் அனைவரும் பயனர்கள் என்று அழைக்கப்படுகின்றனர். இவர்கள் தவிர பயனர் நிர்வாகி, தானியங்கிகள், பயனர் அதிகாரி போன்ற சில உயர்நிலையிலான பயனர்களும் உள்ளனர். இவர்களுக்கு கட்டுரை தொகுப்பில் பயனர்களைக் காட்டிலும் கூடுதலாக சில உரிமைகள் அளிக்கப்பட்டுள்ளன. விக்கிப்பீடியா அமைப்பு எந்தப் பயனர்களுக்கும் பணம் மற்றும் பிற பயன்கள் என்று ஏதும் அளிப்பதில்லை. இருப்பினும் அனைத்து மொழி விக்கிப்பீடியாக்களிலும் தன்னார்வத்துடன் பல லட்சக்கணக்கான பயனர்கள் பதிவு செய்து கொண்டு செயல்பட்டு வருகின்றனர். தமிழ் விக்கிப்பீடியாவில் சுமார் பதினைந்தாயிரம் பயனர்கள் வரை பதிவு செய்துள்ளனர். இவர்களில் சுமார் ஐம்பது பயனர்கள் மட்டும் தொடர்ந்து பங்களித்து வருகின்றனர்.

ஐந்து தூண்கள் கொள்கை

விக்கிப்பீடியாவில்

1. கலைக்களஞ்சியம்,
2. நடுநிலைமை,
3. இலவசக் கலைக்களஞ்சியம்,
4. நன்னடத்தை,
5. கொள்கை மாற்றங்கள்

என்கிற ஐந்து வழியிலான பொதுக் கொள்கைகள் கடைப்பிடிக்கப்படுகின்றன. இவை விக்கிப்பீடியாவின் ஐந்து தூண்கள் என்று சொல்லப்படுகின்றன. தமிழ் விக்கிப்பீடியாவிலும் இக்கொள்கை கடைப்பிடிக்கப்படுகிறது. இதன்படி தமிழ் விக்கிப்பீடியாவில் உருவாக்கப்படும் கட்டுரைகள் பாகுபாடில்லாமல் நடுநிலைமையுடன் பிற மொழிச் சொற்கள் கலப்பின்றி இருக்க வலியுறுத்தப்படுகிறது. கட்டுரைகளில் பிற பயனர்கள் மாற்றம் செய்யும் பொழுது அதை ஏற்றுக் கொள்ள வேண்டும். ஆலமரத்தடி என்கிற கலந்துரையாடல் பக்கத்தின் வழியாக பயனர்களால் பரிந்துரைக்கப்படும் கொள்கைகள் தமிழ் விக்கிப்பீடியாவில் நடைமுறைப்படுத்தப்படுகின்றன. இதுபோல் விக்கிப்பீடியாவின் பெயருக்கு களங்கம் ஏற்படும் நிலையைத் தவிர்க்க சில வழிகாட்டல்களை முன்வைத்து அதைப் பின்பற்றவும் வேண்டப்படுகிறது.

தமிழ் விக்கிப்பீடியாவில் பங்களிப்புகள்

தமிழ் விக்கிப்பீடியா இணையத்தில் இருக்கும் ஒரு இலவசக் கலைக்களஞ்சியம். இதில் உலகின் எப்பகுதியில் இருப்பவரும் தமிழ் விக்கிப்பீடியாவின் கொள்கைகளுக்கும், வழிகாட்டலுக்கும் உட்பட்டு பங்களிப்புகளைச் செய்ய முடியும். தமிழ் விக்கிப்பீடியாவிற்குள் ஏதாவது ஒரு தகவலைத் தேடி பார்வையாளராக வந்தவர்கள் யாரும் அந்தப் பக்கத்தில் ஏதாவது மாற்றம் செய்ய வேண்டும் அல்லது

புதிதாக ஏதாவது சேர்க்க வேண்டும் என்கிற நிலையில் பங்களிக்க முடியும். இதே போல் தமிழ் விக்கிப்பீடியாவின் பயனராகப் பதிவு செய்து கொண்டு புதிய கட்டுரைகள் உருவாக்கம், ஏற்கனவே உள்ள கட்டுரைகளில் தேவையான மாற்றங்கள், தகவல் மற்றும் படம் சேர்க்கைகள் போன்றவற்றை செய்து பங்களிக்க முடியும்.

கட்டுரை உள்ளீடு செய்தல்

தமிழ் விக்கிப்பீடியாவின் ஒவ்வொரு பக்கத்திலும் இடது புறத்தில் இருக்கும் தேடல் என்கிற தலைப்பின் கீழ் காலியாக இருக்கும் பெட்டியில் கட்டுரைக்கான தலைப்பை உள்ளீடு செய்து செல் அல்லது தேடு என்ற பொத்தானைச் சொடுக்கினால், தமிழ் விக்கிப்பீடியாவில் அந்தத் தலைப்பில் கட்டுரை முன்பே இடம் பெற்றிருந்தால் அந்தப் பக்கம் திறக்கும். அந்தத் தலைப்பில் கட்டுரை எதுவும் இல்லாத நிலையில், உள்ளீடு செய்த தலைப்பு சிகப்பு நிறத்தில் தெரிவதுடன் "இந்தத் தலைப்பில் கட்டுரைகள் எதுவுமில்லை, இதை நீங்கள் உருவாக்கலாம்" என்கிற செய்தியும் கிடைக்கிறது. சிகப்பு நிறத்தில் தெரியும் தலைப்பைச் சொடுக்கினால் அந்தத் தலைப்பிற்கான புதிய தொகுத்தல் பக்கம் திறக்கிறது. இந்தப் பக்கத்தில் தமிழ் விக்கிப்பீடியாவிற்கு ஏற்ற முறையில் தமிழில் தட்டச்சு செய்யலாம் அல்லது முன்பே தட்டச்சு செய்யப்பட்ட கட்டுரையை அப்படியே பிரதி (Copy) செய்து உருவாக்க விரும்பும் கட்டுரைக்கான தொகுப்புப் பக்கத்தில் (Paste) ஒட்டி விடலாம்.

தொகுப்புப் பக்கத்தில் தட்டச்சு செய்து முடித்த பின்பு அதன் கீழ் உள்ள "முன் தோற்றம் காட்டு" என்கிற பொத்தானைச் சொடுக்கிக் கட்டுரை இடம் பெறும் தோற்றம் காணலாம். அவை சரியாய் இருக்கும் நிலையில் "பக்கத்தைச் சேமிக்கவும்" என்கிற பொத்தானைச் சொடுக்கினால் தமிழ் விக்கிப்பீடியாவில் அந்தக் கட்டுரை இடம் பெற்று விடும்.

விக்கிக் குறியீடுகள்

தமிழ் விக்கிப்பீடியாவில் கட்டுரைகள் அமைக்கப்படும் போது அதன் அமைப்பில் எளிமையான விக்கிக் குறிகள் சில பயன்படுத்தப்படுகிறது. இந்தக் குறியீடுகள் கட்டுரை தனித்துத் தெரியவும், விக்கிப்பீடியாவின் பிற பக்கங்களுக்குச் செல்ல இணைப்பாகவும், பிற தளங்களுக்குச் செல்ல இணைப்பாகவும் பயன்படுத்தப்படுகிறது. இதில் கீழ்காணும் சில முக்கியக் குறியீடுகளை மட்டும் காணலாம்.

அ.) தலைப்புகள்

கட்டுரையினுள் உள்ள தலைப்புகளுக்கு = எனும் குறியீடு பயன்படுத்தப்படுகிறது. முதன்மைத் தலைப்புகளுக்கு ==முதன்மைத் தலைப்பு== என்று இரு முறையும், துணைத்தலைப்புகளுக்கு ===துணைத்தலைப்பு=== என்றும், அடுத்து வரும் உள்தலைப்புகளுக்கு ஏற்ப = குறியீட்டின் எண்ணிக்கை இருபுறமும் அதிகமாகிறது. இந்தத் தலைப்புகள் நான்கிற்கு அதிகமாகும் போது கட்டுரையின் மேல் பக்கத்தில் நாம் அளித்த தலைப்புகளைக் கொண்டு தானாகவே ஒரு பொருளடக்கப் பெட்டி உருவாகி விடுகிறது. இந்த பொருளடக்கப் பெட்டியைக் காட்டவும் மறைத்துக் கொள்ளவும் வசதி உள்ளது.

ஆ.) எழுத்து அமைப்பு

கட்டுரையில் சில இடங்களில் குறிப்பிட்ட சொற்கள் தனித்துத் தெரிய ஒற்றை மேற்கோள் குறி பயன்படுத்தப்படுகிறது. உதாரணமாக, "சொல்" என்று இருபுறமும் இரண்டு ஒற்றை மேற்கோள்குறியை இரு புறமும் தட்டச்சு செய்தால் சொல் என்று சாய்வெழுத்தாகவும், மூன்று ஒற்றை மேற்கோள்குறியை இரு புறமும் தட்டச்சு செய்தால் சொல் என்று தடித்த எழுத்தாகவும், ஐந்து ஒற்றை மேற்கோள்குறியை இரு புறமும் தட்டச்சு செய்தால் சொல் என்று தடித்த எழுத்தாகவும், சாய்ந்த எழுத்தாகவும் தெரியும்.

இ.) இணைப்புகள்

கட்டுரையின் உள்ளே இடம் பெற்றிருக்கும் ஒரு சொல்லில் உள்ள கட்டுரைக்குச் செல்ல சதுர அடைப்புக்குறி பயன்படுத்தப்படுகிறது. கட்டுரை இப்பெயரில் இருக்கலாம் என்று கருதும் சொல்லின் இருபுறமும் இரு சதுர அடைப்புக்குறிகளை இட வேண்டும். உதாரணமாக, கட்டுரையிலுள்ள கோயம்புத்தூர் என்ற சொல்லின் இருபுறமும் சதுர அடைப்புக் குறிகளைப் பயன்படுத்தி [[கோயம்புத்தூர்]] என்று தட்டச்சு செய்து விட்டால் தமிழ் விக்கிப்பீடியாவிலுள்ள கோயம்புத்தூர் என்ற கட்டுரைப் பக்கத்திற்கு அங்கிருந்து இணைப்பு செய்யப்பட்டு விடுகிறது. கோயம்புத்தூர் என்ற பெயரில் கட்டுரை இருந்தால் நீலநிறத்திலும் கட்டுரை இல்லாவிடில் சிகப்பு நிறத்திலும் தெரியும். நீலநிறமாக இருக்கும் பெயரில் சொடுக்கினால் அப்பெயரிலான கட்டுரைக்கு நேரடியாகச் சென்றுவிடுகிறது. சிகப்பு நிறமாக இருக்கும் பெயரில் சொடுக்கினால் அந்தப்பக்கத்தை உருவாக்கச் சொல்லி வேண்டுவதுடன் அதற்கான தொகுப்புப் பக்கமும் திறக்கிறது. இது உள் இணைப்புகள் எனப்படுகிறது.

இதுபோல் கட்டுரையில் தேவையான இடங்களில் பிற இணைய தளங்களுக்கு இணைப்பு செய்ய ஒரு சதுர அடைப்புக் குறியுடன் அந்த இணைய தள முகவரியை அளித்து நாம் அளிக்கும் பெயரையும் சேர்க்கலாம். உதாரணமாக உலகத்தமிழ் செம்மொழி மாநாடு தளத்திற்கு இணைப்பு செய்ய இணைய முகவரியிட்டு சிறிது இடைவெளி விட்டு உலகத் தமிழ் செம்மொழி மாநாடு என்று தட்டச்சு செய்து இருபுறமும் ஒரு சதுர அடைப்புக் குறியிட்டு விடவேண்டும். அதாவது, [<http://www.ulakathamizhchemmozhi.org/> உலகத்தமிழ் செம்மொழி மாநாடு] என்று தட்டச்சு செய்தால் உலகத்தமிழ் செம்மொழி மாநாடு என்று நீலநிறத்தில் தனியே தெரிவதுடன் அதைச் சொடுக்கினால் அந்தத்தளத்திற்கு நேரடியாகச் சென்று விடுகிறது. இது வெளி இணைப்புகள் எனப்படுகிறது

ஈ.) அடிக்குறிப்புகள் மற்றும் மேற்கோள்கள்

கட்டுரையில் மேற்கோள்கள் காட்ட வேண்டிய இடங்களில் <ref> என்று தட்டச்சு செய்து அடிக்குறிப்புத் தகவல்களை குறிப்பிட்டு அதன்பின்பு </ref> என்று தட்டச்சு செய்து விடலாம். கடைசியாக தலைப்புகளில் ஒன்றாக மேற்கோள்கள் என்று குறிப்பிட்டு அதன் கீழாக <reference/> அல்லது {{Reflist}} என்று தட்டச்சு செய்து விட்டால் வரிசையாக எண்ணிடப்பட்டு கொடுக்கப்பட்ட அனைத்து அடிக்குறிப்புகளும் மேற்கோள்கள் என்ற தலைப்பின் கீழ் ↑ என்ற குறியீட்டுடன் தனியாகத் தரப்பட்டு விடுகின்றன.

உ.) பிற குறியீடுகள்

இதுபோல் கட்டுரைகளில் புள்ளியிடலுக்கு * என்ற குறியீடும், எண்ணிடலுக்கு # என்ற குறியீடும், வரி தள்ளலுக்கு : என்ற குறியீடும் பயன்படுத்தப்படுகின்றன. இதுபோல் கட்டுரைகளுக்குத் தேவையான செயல்பாடுகளுக்கு ஏற்ற எளிமையான விக்கிக் குறியீடுகளை அளித்து கட்டுரையைச் சிறப்பாக்க விக்கிப்பீடியாவின் தனி மென்பொருள் உதவுகிறது.

படிமங்கள்

கட்டுரைக்குத் தேவையான படிமங்களையும் எளிதில் சேர்க்க முடியும். தமிழ் விக்கிப்பீடியாவின் அனைத்துப் பக்கங்களிலும் இடது புறமுள்ள கோப்பைப் பதிவேற்று என்கிற இடத்தில் சொடுக்கினால் அதற்கான பக்கம் திறக்கிறது. இதில் விக்கிப்பீடியாவின் கட்டுப்பாடுகள் மற்றும் காப்புரிமை குறித்த சில விபரங்களைத் தெரிந்து கொண்டு, தேவையான விபரங்களைப் பதிவு செய்து, கட்டுரைக்குத் தேவையான படிமங்களைக் அதற்கான பெயர்களில் பதிவேற்றி விடலாம். இதில் விக்கிப்பீடியாவின் விதிமுறைகளுக்குப் பொருந்தாத படிமங்கள் குறிப்பிட்ட கால இடைவெளிக்குப் பின்பு நீக்கப்பட்டுவிடும். பதிவேற்றம் செய்யப்பட்ட படிமங்களை தமிழ் விக்கிப்பீடியாவின் கட்டுரைகளில் தேவையான இடங்களில் [[படிமம்: படிமத்தின் பெயர்.jpg]] என்று தட்டச்சு செய்தால் கட்டுரையில் படிமம் இடம் பெற்று விடுகிறது. இந்தப் படிமத்தை

தேவைக்கேற்ப வலது, இடது அல்லது மத்தியப்பகுதியில் இணைக்கவும், படிமங்களின் கீழ் குறிப்புகளிடவும், படிமங்களை தேவைப்பட்டால் பார்வையிடவும், படிமங்களுக்கு தனியே இணைப்பு செய்து கொள்ளவும் சில குறியீடுகள் சேர்த்துப் பயன்படுத்தப்படுகின்றன.

வார்ப்புருக்கள்

தமிழ் விக்கிப்பீடியாவில் தயார் நிலையிலுள்ள சில வார்ப்புருக்கள் பட்டியலிடப்பட்டுள்ளன. இந்த வார்ப்புருக்கள் பட்டியலிலிருந்து நமக்குத் தேவையான வார்ப்புருக்களைக் கட்டுரையில் தேவையான இடங்களில் பயன்படுத்திக் கொள்ளமுடியும். இதற்கு {{வார்ப்புருவின் பெயர்}} என்று தட்டச்சு செய்து விட்டால் கட்டுரையில் குறிப்பிட்ட வார்ப்புரு இடம் பெற்றுவிடும். இந்த வார்ப்புருக்களில் குறிப்பிட்ட இடங்களில் தேவைக்கேற்ப மாற்றங்கள் செய்தும் பயன்படுத்திக் கொள்ள முடியும்.

அட்டவணைகள்

தமிழ் விக்கிப்பீடியாவில் அட்டவணைகளை உருவாக்கிக் கொள்ள சில எளிய வழிமுறைகள் உள்ளன. இந்த வழிமுறைகளைப் பயன்படுத்திக் கட்டுரையில் தேவையான இடங்களில் தேவையான அட்டவணைகளை இடம் பெறச் செய்ய முடியும்.

பகுப்புகள்

தமிழ் விக்கிப்பீடியாவில் தமிழ், பண்பாடு, வரலாறு, அறிவியல், கணிதம், தொழில்நுட்பம், புவியியல், சமூகம், நபர்கள் என்று முக்கிய பகுப்புகளின் கீழ் கட்டுரைகளைக் கொண்டு வரமுடியும். இதற்கு கட்டுரையின் கீழ்ப்பகுதியில் [[பகுப்பு: தமிழ்]] அல்லது குறிப்பிட்ட பகுப்புகளின் கீழ் குறிப்பிடப்பட்டால் அந்தக் கட்டுரையின் தலைப்பு குறிப்பிட்ட பகுப்பின் கீழ் இடம் பெற்றுவிடும். இந்த முக்கியப் பகுப்பின் கீழ் அடங்கிய உள் பகுப்புகளின் கீழும் உள்ளீடு செய்ய முடியும்.

கட்டுரை செயலிகள்

தமிழ் விக்கிப்பீடியாவில் இடம் பெறும் ஒவ்வொரு கட்டுரைக்கும் மேல்பக்கத்தில் கட்டுரை, உரையாடல், தொகு, வரலாறு என்று சில செயலிகள் இடம் பெறுகிறது. கட்டுரைப் பக்கத்தில் சொடுக்கினால் கட்டுரையும், உரையாடல் பொத்தானைச் சொடுக்கினால் அந்தக் கட்டுரை குறித்த கருத்துக்கள் பதிவு செய்யப்பட்டிருக்கும் பகுதியும் தெரிகின்றன. தொகு எனும் பொத்தானைச் சொடுக்கினால் கட்டுரையில் தேவையான இடங்களில் மாற்றம் செய்யவும், தகவல்களைச் சேர்க்கவும் முடியும். வரலாறு எனும் பொத்தானைச் சொடுக்கினால் அந்த கட்டுரையில் செய்யப்பட்ட அனைத்து மாற்றங்களும் பேச்சு, உரையாடல், நாள், நேரம், பைட்ஸ் அளவு, மாற்றம் செய்த பயனர் அல்லது பயன்படுத்தியவரது இணைய விதிமுறை இலக்கம் (I.P.Number) போன்ற விபரங்கள் இடம் பெறுகின்றன. இதுபோல் அண்மைய மாற்றங்கள் எனும் தலைப்பிலான தனிப்பக்கத்திலும் தமிழ் விக்கிப்பீடியாவில் செய்யப்படும் அனைத்து மாற்றங்களும் உடனுக்குடன் பதிவு செய்யப்பட்டு விடுகின்றன.

தரம் நிர்ணயம்

தமிழ் விக்கிப்பீடியாவில் இடம் பெற்றிருக்கும் சில கட்டுரைகள் தரத்தின் அடிப்படையில் சிறப்பு, மிகவும் நல்லது, நல்லது, துவக்கம், குறுங்கட்டுரை என்கிற ஐந்து பிரிவுகளாக வகைப்படுத்தப்படுகிறது. இதுபோல் கட்டுரையின் முக்கியத்துவம் கருதி மிக உயர்வு, உயர்வு, நடுநிலை, தாழ்வு என்றும் வகைப்படுத்தப்படுகிறது. இந்த வகைப்பாடுகளுக்கு குறிப்பிட்ட சில நிறங்கள் மட்டுமே அடையாளமாக கொள்ளப்படுகின்றன. இவற்றின் அடிப்படையில் சில முக்கியக் கட்டுரைகள் தேர்வு செய்யப்பட்டு சிறப்புக் கட்டுரைகளாக்கப்படுவதுடன் முதற்பக்கத்தில் குறிப்புகள் வெளியிடப்பட்டு இதற்கான இணைப்பும் தரப்படுகின்றன. மிக முக்கியமான மற்றும் முழுமையான கட்டுரைகள் பிற பயனர்களால் மாற்றம் செய்ய

முடியாதபடி பயனர் நிர்வாகிகளால் பூட்டிடப்பட்டு விடுகின்றன. இதனால் முழுமையடைந்த கட்டுரைகளில் தேவையற்ற மாற்றங்கள் செய்யப்படுவது தடுக்கப்படுகின்றன.

குறைபாடுகள்

விக்கிப்பீடியாவில் எவரும், எப்போது வேண்டுமானாலும் பங்களிக்க முடியும் என்கிற பொதுவான நிலை உள்ளது. இதனால் சில சமயம் இங்குள்ள கட்டுரைகளைப் பார்வையிடும் சில தவறான நோக்கமுடைய, விஷமத்தனமுடையவர்களால் சில கட்டுரைகள் தவறுதலான மாற்றத்துக்கு உள்ளாகிறது. இதை அவ்வப்போது கவனித்து வரும் பயனர் நிர்வாகிகள் முன்பிருந்த நிலைக்கு மீட்டெடுக்கின்றனர். தவறும் நிலையில் கட்டுரையில் பதிவான தவறான கருத்துக்கள் கண்டுபிடிக்கப்படும் வரை தொடர்கிறது. தமிழ் விக்கிப்பீடியாவில் பதிவு செய்யப்படும் பல கட்டுரைகள் குறுங்கட்டுரைகளாகத் துணுக்குத் தகவல்களைப் போல் இருக்கின்றன என்கிற குறைபாடும் உள்ளது. இவை குறுங்கட்டுரைகள் எனும் தனிப் பகுப்புகளின் கீழ் இருக்கின்றன. இவைகளைப் பார்வையிடும் பயனர்கள் விரிவாக்கம் செய்யும் வரை இவை குறுங்கட்டுரைகளாகவே தொடர்கின்றன.

முடிவுரை

இது போன்ற ஒரு சில குறைபாடுகள் இருக்கின்ற நிலையிலும் தமிழ் விக்கிப்பீடியா தன்னார்வப் பயனர்கள் பலரின் கூட்டு முயற்சியோடும் அவர்களின் பங்களிப்போடும் மிக வேகமாக வளர்ந்து வருகிறது. உலகின் 267 மொழிகளில் இருக்கும் விக்கிப்பீடியாக்களில் தமிழ் விக்கிப்பீடியா 22,000-க்கும் அதிகமான கட்டுரைகளுடன் 67 வது நிலையில் இருக்கிறது. இந்நிலையில் தமிழ்நாடு அரசு, தமிழ் விக்கிப்பீடியாவின் நிலையை உயர்த்த உலகத் தமிழ் செம்மொழி மாநாட்டுடன் இணைந்த தமிழ் இணைய மாநாட்டின் ஒரு நிகழ்வாக கல்லூரி மாணவர்களுக்கு விக்கிப்பீடியா தகவல் பக்கங்கள் போட்டி ஒன்றை நடத்த முன் வந்துள்ளது. இதன் மூலம் மாணவ சமுதாயத்திலிருந்து தமிழ் விக்கிப்பீடியாவிற்கு பங்களிக்கும் புதிய பயனர்களை உருவாக்க உதவியுள்ளது. மேலும் இந்தப் போட்டிக்கு வரும் கட்டுரைகளில் தேர்வு செய்யப்படும் கட்டுரைகள் அனைத்தும் தமிழ் விக்கிப்பீடியாவில் பதிவேற்றம் செய்யப்பட உள்ளன. இதனால் தமிழ் விக்கிப்பீடியா இந்திய மொழிகளில் முதலிடத்தையும், உலக மொழிகளில் சிறிது முன்னேற்றத்தையும் அடையும். உலகம் முழுவதுமுள்ள தமிழர்கள் தங்களைத் தமிழ் விக்கிப்பீடியாவின் பயனர்களாகப் பதிவு செய்து கொண்டு தங்களுக்குத் தெரிந்த துறையிலான தகவல்களை பதிவேற்றம் செய்து தமிழ் விக்கிப்பீடியா எனும் கலைக்களஞ்சியத் தொகுப்பில் பங்கேற்க முன்வர வேண்டும்.

தமிழில் கணினிவழிச் சொல்லடைவு

இல. சுந்தரம்

MA (Tamil), M.Sc (I.T.), M.C.A., M.Phil.,(Tamil)

கணினித் திட்ட அமைப்பாளர்,

கலைஞர் வளர்தமிழ் மையம்,

பாரதிதாசன் பல்கலைக்கழகம்,

திருச்சிராப்பள்ளி.

e-mail : sundarbaskar@gmail.com

முன்னுரை

தமிழ் மொழியானது வளர்ந்துகொண்டேவரும் அறிவியல் மற்றும் தொழில்நுட்பத்திற்கு ஈடாகத் தன்னையும் வளர்த்துக்கொண்டே வருகின்றது. இன்றைய கணினி, இணைய, கைபேசி தொழில்நுட்ப உலகில் பல்வேறுபட்ட மென்மங்கள் தமிழ்மொழிக்கென உருவாக்கப்படுகின்றன.

கணினிமொழியியல் கோட்பாடுகளை வைத்து நம் மொழியின் அமைப்பை நிரலிகளாக அமைத்து தமிழ்மொழியின் தேவையை நிறைவுசெய்ய வேண்டும். அவ்வகையில் சொல்லடைவு, தொடரடைவு என்பது என்ன? தமிழில் இதுவரை கணினிவழிச் சொல்லடைவு உருவாக்குவதற்கு மேற்கொள்ளப்பட்ட முயற்சிகள் மற்றும் அதன் சுருக்க வரலாறு, பிறமொழிகளில் சொல்லடைவு மென்ம உருவாக்கத்தின் வளர்ச்சி, தமிழில் இதன் நிலை, தமிழில் நடையியல் ஆய்வுக்கு இது எவ்வாறு உதவும், தமிழில் சொல்லடைவு மென்மம் உருவாக்குவதால் விளையும் பயன், இந்தச் சொல்லடைவு மென்பொருளை உருவாக்கும்போது ஏற்படுகின்ற சிக்கல்கள் போன்றவற்றை ஆராய்வதாக இக் கட்டுரை அமைகிறது.

தற்போது சொல்லாளர்,தானியங்கு சொற்பிழைதிருத்தி, சந்திப்பிழைதிருத்தி, எழுத்து-பேச்சு மாற்றி, தானியங்கு பேச்சு அறிவான், ஒளிவழி எழுத்தறிவான், இணையம் தொடர்பான மென்மங்கள் எனப் பல்வேறு நிலைகளில் தமிழ் மென்மங்கள் உருவாக்கப்படுகின்றன. அவ்வகையில் தமிழ் நூல்களுக்குக் கணினிவழிச் சொல்லடைவு மென்மம் தயாரிப்பதிலும் கவனம் செலுத்தவேண்டிய கட்டாயம் ஏற்பட்டுள்ளது. யூனிகோடு என்ற உலகமயமாக்கலில் எந்த ஒரு முயற்சியும் அனைவருக்கும் பயனளிக்கும் வகையிலேயே அமைகின்றது.

தமிழில் 18-ஆம் நூற்றாண்டிற்குப் பிறகும் மொழியியல் தமிழில் கால்கோலிட்டதற்குப் பிறகும் சிற்சில மொழிக்கூறுகளைத் தனிப்படப் பிரித்துத் தொகுத்து அடைவுபடுத்த முயன்றனர். அவ்வகையில் மனித உழைப்பை மட்டுமே வைத்து அகராதிகளைத் தொகுத்தனர், பின்னர் இலக்கியப் படைப்பு ஒவ்வொன்றிற்கும் பயன்பாடு கருதிப் பல்வேறு அடைவுகளை உருவாக்கினர். தற்போது கணினி மொழியியல் வளர்ச்சியின் பயனாகப் பல்வேறு நுண்மொழிக்கூறுகளும் கண்டறியப்பட்டு அதற்கென மென்மங்கள் உருவாக்கத் தொடங்கியுள்ளனர். இன்றைய தகவல் உலகில், எதை? எப்படி வேண்டுமானாலும், தரவுகளை, தகவல்களைத் தொகுத்துப் புதியதாக மாற்றியமைக்கலாம். இவ்வகையில் பல்வேறு நுண்மொழிக்கூறுகளும் கண்டறியப்படுகின்றன.

மொழியின் தேவை

தொடரடைவினை நாம் கைவேலையாகச் செய்யும்போது அது மிகவும் கடினமானதொரு பணி என்பதை உணர முடியும். ஆகவேதான் பொதுவாகப் பெரிய இலக்கியப் படைப்புகளுக்கு மட்டுமே தொடரடைவுகள் செய்யப்படுகின்றன. ஷேக்ஸ்பியர் நூல்கள், பைபிள், திருக்குறள் போன்றவற்றிற்குத் தொடரடைவு செய்யப்பட்டுள்ளது. எனவே இதிலிருந்து அனைத்துப் படைப்புகளுக்கும் தொடரடைவு

தேவையென்றாலும்கூட அதனைக் கைப்பணியாகச் செய்வது மிகவும் கடினமானது என்பதால் முக்கியமாக்கக் கருதப்பட்ட நூல்களுக்கே தொடரடைவு செய்யப்பட்டுள்ள நிலையை அறிய முடிகிறது.

சொல்லடைவு(Index) - தொடரடைவு(Concordance) - பொருளடைவு(Subject Index)

ஒரு சொல் ஒரு நூலில் எந்தெந்த இடங்களில் வருகிறது என்பதும் முக்கியக் கலைச்சொற்களும் நூலின் பின்பகுதியில் கொடுக்கப்படும். வெறும் சொல்லும் அது வருமிடமும் கொடுக்கப்பட்டால் அது சொல்லடைவு, அந்தச் சொல் வரும் தொடரை அப்படியே எடுத்துக் கொடுப்பது தொடரடைவாகும்; அந்தத் தொடர்களில் எந்தெந்தப் பொருள்களில் வருகிறது என்பதையும் அதன் இலக்கணத் தன்மை போன்றவற்றைச் சேர்த்துக் கொடுப்பதும் தொடரடைவாகும். பொருளடைவு என்பது ஒரு படைப்பில் ஒரு சொல் எங்கெங்கு என்னென்ன பொருளில் வருகிறது என்பதை ஆராய்ந்து அடைவுபடுத்துவதாகும்.

இந்த மூன்றும் ஒன்றுக்கொன்று தொடர்புடையவை இதை முதல்நிலை(First Stage), இரண்டாம் நிலை, மூன்றாம் நிலை எனவும் கூறலாம். பொருளடைவினைக் கணினிவழி உருவாக்குவது மிகக் கடினம். ஏனெனில் அவற்றின் பொருண்மையை மனித உழைப்பால்தான் தீர்மானிக்க முடியும்.

சொல்லடைவு, தொடரடைவு, பொருளடைவு என்பது என்ன? என்பதற்கு முழுமையான வரையறை உருவாக்கப்படவில்லை எனலாம். ஏனெனில் அது யாருக்காக உருவாக்கப்படுகிறது என்பதைப்பொறுத்து மாறுபடுகிறது. மேற்குறிப்பிட்டது ஒரு சிறிய அடிப்படை.

சொல்லடைவு என்பது ஒவ்வொரு படைப்பு அல்லது நூலுக்கு கடைசிப் பகுதியில் அந்தப் படைப்பில் பயன்படுத்தப்பட்டுள்ள சொற்களை அகர வரிசைப்படுத்தி அவற்றின் பயிற்றிடத்தைப்(பக்க எண், அல்லது பாடல் எண்) பட்டியலிடுவதாகும். ஒரு சொல் அந்தப் படைப்பில் எந்தெந்த இடங்களில் வருகின்றது என்பதைச் சுட்டிக்காட்டுவதால் இதனைச் 'சுட்டி' என்றும் அழைக்கின்றனர்.

சொல்லடைவு உருவாக்குவதில் பல நிலையுண்டு, படைப்பில் உள்ள எல்லாச் சொற்களையும் அடைவுபடுத்துவது, அதில் காணப்படும் அருஞ்சொற்களுக்கு மட்டும் அடைவுகொடுப்பது, பெயர், வினை ஆகியவற்றிற்கு மட்டும் கொடுத்து வேற்றுமை உருபு, சாரியை, பெயரடை, வினையடை போன்றவற்றை விட்டுவிடுவது எனப் பல்வேறு நிலைகளில் உருவாக்கப்படுகின்றது.

ஒரு சொல்லுக்கு நாம் விருப்பத்துக்கேற்ப பொருள் காணாமல், இலக்கியத்தில் எந்தெந்த இடங்களில் அச்சொல் வருகிறது, அதற்கு அந்த இடத்தில் என்ன பொருள், அதற்குப் பழைய உரையாசிரியரின் ஆதாரம் உண்டா, காலப்போக்கில் அச் சொல்லின் பொருள் எவ்வாறு மாற்றம் அடைந்துள்ளது என்பவற்றையெல்லாம் ஆராய்ந்து அடைவு தயாரிப்பதையே பொருளடைவு என்று கூறுவர். சொற்களின் படிநிலை அமைப்பில் எங்கு வரும் என்பதன் அடிப்படையில்தான் சொற்களின் பொருள்களைப் பெறமுடியும்.

தமிழில் மனித உழைப்பால் கணினி உதவியின்றிப் பல்வேறு சொல்லடைவுகள் உருவாக்கப்பட்டன அவற்றுள் சில, திருக்குறட் சொல்லடைவு(1952);சாமி. வேலாயுதம், பழந்தமிழ்ச் சொல்லடைவு(1957);நீ.கந்தசாமி, தொல்காப்பியச் சொல்லடைவு(1968), புறநானூறு சொல்லடைவு; வ.அய் சுப்பிரமணியம், தொல்காப்பியச் சிறப்பகராதி(2000); ப.வே.நாகராசன், த.விஷ்ணுகுமாரன் திருவனந்தபுரம் பன்னாட்டுத் திராவிட மொழியியற் கழகம், சங்க இலக்கியச் சொல்லடைவு(2001);பெ.மாதையன்.

1. 1985;Computer Analysis of Tirukural;S.Baskaran, Cellamuthu;Tamil University,
2. 1993;A word index of old Tamil Cankam literature;Thomas Lehman and Thomas Malten;Institute of Asian Studies, Chennai போன்றவை கணினிவழிச் சொல்லடைவு நூல்கள்.

தமிழ்ப் பல்கலைக்கழகத்தில் 1986-ல் டாக்டர் ச.பாஸ்கரன் அவர்களின் குழுவினரால் பாவலரேறு பாலசுந்தரம் அவர்களின் வழிகாட்டுதலில் திருவாசகம், சங்க இலக்கியம் ஆகியவற்றிற்குக் கணினிவழிச் சொல்லடைவு உருவாக்கப்பட்டது, இதில் பொருளடைவும் சேர்க்கப்பட்டுள்ளது.

சொற்களஞ்சியம்

சொற்களஞ்சியம் என்பது ஓர் அகராதியை அடிப்படையாகக்கொண்டு அதில் உள்ள சொற்களைப் பொருள் அடிப்படையிலும் அவற்றின் பொருண்மை உறவு அடிப்படையிலும் தொடர்புபடுத்தும் நோக்கில் அமைப்பதாகும். சொற்பொருளைத் தெரிந்துகொள்ள அகராதிகளும் முக்கியச் செய்திகளையும் கோட்பாடுகளையும் விளங்கிக்கொள்ள கலைக்களஞ்சியங்களும் தகவல் களஞ்சியங்களும் பெரிதும் துணைநிற்கின்றன. இவ்வகையில் தற்கால வளர்ச்சி நிலைகளாக மின் அகராதிகளும் மின்சொற்களஞ்சியங்களும் தமிழில் உருவாக்கப்படுகின்றன.

இருமொழி, மும்மொழி, அடுக்குமொழி, மரபுத்தொடர், பழமொழி, ஆட்சிச்சொல், அரும்பொருள் விளக்க, அறிஞர் தமிழ், இலக்கியச்சொல், இலக்கண, எதுகை, ஒலிக்குறிப்பு, கலைச்சொல், சிறப்புப் பெயர், தமிழ்ப்புலவர், அருந்தொடர், நுட்பச்சொல் போன்ற அகராதிகளும் கலை, தகவல், குழந்தைகள், சொற்களஞ்சியம் போன்ற களஞ்சியங்களும் பதமஞ்சரி, நூல்விவர அட்டவணை போன்றவை தமிழில் 1950களுக்குப் பிறகு பல்வேறு நிலைகளில் பல்வேறு வகைப்பாடுகளில் உருவாக்கப்பட்டுள்ளன. இவையெல்லாம் கணினிமயமாக்கப்படவேண்டும். தற்காலத் தமிழ்ச் சொற்களஞ்சியம்(2001);தமிழ்ப் பல்கலைக்கழகத்தின் வாயிலாக வெளியிடப்பட்டுள்ளது. தமிழ் மின்சொற்களஞ்சியம்(2006); ச.இராசேந்திரன், ச.பாஸ்கரன் ஆகியோரால் எனும் நூல் வெளியிடப்பட்டுள்ளது.

நடையியல் ஆய்வு

கணினி வழியாக ஒரு படைப்பின் நடையியலையும் அதன் கட்டமைப்பினையும் பல்வேறு பரிமாணங்களில் ஆய்வு செய்வது பனுவல் ஆய்வு(Text Analysis) எனப்படும். பேச்சு நடையும் எழுத்து நடையும் ஒருவரை ஒருவர் வேறுபடுத்திக்காட்ட உதவுகிறது. இவ்வகையில் ஒவ்வொருவருடைய எழுத்து நடையை அடையாளம் காணவும் அகராதி தொகுக்கும் நோக்கத்திற்கும் (Lexical Analysis) தொடரியல் (Syntactic Analysis), சொற்பொருள் (Semantic Analysis) பகுப்பாய்வுகளுக்கும் இந்த நடையியல் ஆய்வு பெரிதும் உதவுகின்றது. இதனை மொழிநடை ஆய்வு (Stylistics Study) எனவும் அழைக்கலாம்.

நடை ஓர் ஆசிரியனின் தனித்தன்மையை வெளிப்படுத்தவல்லது. படைப்பாளனை இனம் கண்டுகொள்ளும் வகையில் தெளிவுபடுத்துவது. இவ்வகையில் சங்க இலக்கியச் சொற்களையும் திருக்குறட் சொற்களையும் ஒப்பிட்டு வள்ளுவர் தமிழுக்கு அறிமுகப்படுத்தியச் சொற்கள் எவை? அவற்றுள் எவையெல்லாம் இன்றைய மொழியில் நிலைக்கொண்டுள்ளன, இவற்றை அறியும் முயற்சியாக மொழி அறக்கட்டளையின் வாயிலாக பா.ரா சுப்பிரமணியன் கட்டுரை எழுதியுள்ளார்.

கணினிவழிச் சொல்லடைவு மென்மம் உருவாக்கம்

சொல் தேடல்(Word Search) நிலையில் வேர்ச்சொல் தேடல்(Root word Search), முழுச்சொல் தேடுதல்(Full Word Search) என்னும் நிலையில் சொற்களை வகைப்படுத்தி தேர்ந்தெடுக்க முடிகிறது. வரிசைப்படுத்தல்(Sorting) நிலையில் சொற்பட்டியலை வருகைமுறைப்படி (Running Type), அகர வரிசைப்படி(Alphabetical Order), நிகழ்வெண்ணிக்கையின்படி(Occurrence) எனப் பலவாறு வகைப்படுத்த முடியும். சொற்களின் வருகைமுறை விகிதத்தையும் கண்டறியமுடியும். கணக்கிடுதல் நிலையில்(Counting) எழுத்து, சொல், தொடர், பத்தி ஆகியவற்றின் எண்ணிக்கையையும் கணக்கிட்டுக்காட்ட முடியும்.

கணினிக்கு நிரல்களின் வாயிலாகச் சொல்லிக்கொடுக்கும்போது அடிப்படை நிலையில் எழுத்துக்களுக்கிடையே இடைவெளி(Space) விடுவதையே சொல்பிரிப்பானாகவும் (Word Split Marker)

நிறுத்தற்குறியை (Full stop) வாக்கியப் பிரிப்பானாகவும் (Sentence Split Marker) பயன்படுத்தமுடியும். இவ்வாறு பயன்படுத்தும்போது பல்வேறுவகையான மொழியமைப்புச் சிக்கல்கள் ஏற்படுகிறது.

உருபன் பகுப்பாய்வு

உருபன் பகுப்பாய்வு (Morphological Parsing) அடிப்படையில் பனுவலைப் பிரித்து சொல்லடைவு உருவாக்கவேண்டிய கட்டாயமும் இங்கு காணப்படுகிறது. இதற்காக உருபன் பகுப்பாய்வி என்னும் மென்மம் உருவாக்கும் பணியில் பேரா. ந. தெய்வசுந்தரம், பேரா. மா. கணேசன் போன்றோர் ஈடுபட்டுள்ளனர். இதில் வெற்றியும் கண்டுள்ளனர் எனலாம்.

கணினிவழிச் சொல்லடைவு சுருக்க வரலாறு

- ❖ திருக்குறளுக்கும் கலித்தொகைக்கும் இணையவழிச் சொல்லடைவினை (தேடுதல்(Search) நிலையில்) தமிழ் இணையப் பல்கலைக்கழகம் உருவாக்கியுள்ளது. அண்ணாமலைப் பல்கலைக்கழக மொழியியல் துறைப் பேராசிரியர் மா.கணேசன் அவர்களால் தமிழ் தகவல்தளத்தைப் பயன்படுத்தி ஒரு குறிப்பிட்ட சொல்லை அடிப்படையாக வைத்து உருவாக்கப்பட்ட சொற்றொடர்கள் போன்றவற்றைக் கண்டறியும் KWIC Concordance, Lemma Extractor உள்ள சொல்லடைவு மென்மம் (Corpus Analysis Tool for Tamil) உருவாக்கப்பட்டுள்ளது.
- ❖ மகிழ்கல்வி நிறுவனத்தின் வாயிலாக திரு. வி. கிருஷ்ணமூர்த்தி அவர்களாலும் இம் முயற்சி மேற்கொள்ளப்பட்டுள்ளது. கிரியா தற்காலத் தமிழ் அகராதியை உருவாக்குவதற்காக அவர்களின் தேவைக்கேற்ப சொல்லடைவு மென்மம் உருவாக்கிப் பயன்படுத்தியுள்ளனர். சென்னைக் கிறித்தவக் கல்லூரி புள்ளியியல் துறையில் திருக்குறளுக்குச் கணினிவழிச் சொல்லடைவு உருவாக்கப்பட்டதாகத் தெரிகிறது.
- ❖ சக்தி ஆபிஸ் மென்மத்தில் Indexing எனும் கருவி(Tool) உருவாக்கப்பட்டுள்ளது. தமிழ்மொழிக்கேற்பச் செயல்படுத்த சென்னைக் கவிகள் நிறுவனத்தால் பல்வேறு முயற்சிகள் மேற்கொள்ளப்பட்டுள்ளது. திரு. கபிலன் அவர்களால் கணியத்தமிழ் நிறுவனமும் இம் முயற்சியில் ஈடுபட்டது. தற்போது செம்மொழித் தமிழாய்வு மத்திய நிறுவனத்தால் சங்க இலக்கியம் மற்றும் மற்ற படைப்புகளுக்குக் கணினிவழிச் சொல்லடைவு உருவாக்கப்பட்டுவருகிறது.

பிற மொழிகளில் சொல்லடைவு

Concordance என்னும் பெயரிலேயே ஆங்கிலத்திற்கு மென்மம் உள்ளது. இந்த மென்மம் ஆஸ்கி ASCII-யில் மட்டுமே செயல்படக்கூடியதாக இருக்கிறது. யூனிகோடு-இல் உள்ளவற்றை செயல்படுத்தமுடியவில்லை. இதுபோன்று இணையத்தில் ஆங்கிலத்திற்குப் பல சொல்லடைவு, தொடரடைவு மென்மங்கள் கிடைக்கின்றன. Simple Concordance எனும் ஆங்கிலத்திலும் எந்தமொழியையும் ஒலிபெயர்ப்பு செய்து பயன்படுத்தும் வகையிலும் நம் தேவைக்கேற்ப அகரவரிசைப்படுத்தும் வசதியுடனும் இந்தச் சொல்லடைவு மென்மம் இணையத்தில் இலவசமாகக் கிடைக்கிறது.

சக்தி ஆபிஸ் என்னும் இந்தி மொழி பதிப்பில் இந்தி மொழிக்குரிய சொல்லடைவுக் கருவி உள்ளது. மலையாள மொழியில் பைபிள் தொடரடைவு மற்றும் அகராதி உருவாக்கப்பட்டுள்ளதாகத் தெரிகிறது. தெலுங்கு மொழியிலும் பைபிளோடு தொடர்புடைய ஹிப்ரு-கிரீக்-தெலுங்கு மினி பைபிள் தொடரடைவு தயாரிக்கப்பட்டுள்ளதாகத் தெரிகிறது.

கணினிவழிச் சொல்லடைவு உருவாக்கத்தில் தமிழ்மொழியமைப்புச் சிக்கல்கள்

சொற்களைப் பிரித்து வரிசைப்படுத்தும்போது சொற்பிரிப்பில் (Word Space, Word Form) பல்வேறு சிக்கல்கள் எழுகின்றன. தமிழ்மொழியைப் பொறுத்தவரை சொற்களை எங்கு பிரிக்க(உடைக்க) வேண்டும்

என்ற கட்டுப்பாடு கிடையாது. தமிழில் சுட்டுப் பெயர்களைக் கொண்டு உருவாக்கப்படுகின்ற சொற்களில் ஒற்று சேர்ந்து இருப்பது ஒரு சொல்லாகவும் ஒற்று இல்லாமல் இருப்பது ஒரு சொல்லாகவும் தனித்தனியாக பிரிக்கப்படுகிறது. உதாரணமாக அந்தக் கடை, அந்த இடம் என்னும் இரண்டு சொற்களில் அந்தக், அந்த என்பது தனித்தனி சொல்லாக வரிசைப்படுத்தப்படும் மேலும் ஒற்று மிகக்கூடிய க, ச, த, ப (அந்தக், அந்தச், அந்தத், அந்தப்) எனும் நான்கும் மிகாமல் வரக்கூடிய அந்த என்ற ஒன்றும் என ஐந்து இடங்களில் இதன் வருகை காணப்படும். இதனால் ஒரே சொல் பட்டியலில் பல இடங்களில் வரக்கூடியதாக இருக்கிறது.

சொற்களைப் பல இடங்களில் பிரித்தும் சேர்த்தும் எழுதுகின்ற வழக்கம் தமிழில் அதிகமாகக் காணப்படுகிறது. அறிந்துகொள்ள என்பதை அறிந்து, கொள்ள என இரண்டு சொற்களாகப் பிரித்து எழுதுகின்றனர். செய்ய வேண்டும், காணவேண்டும் போன்ற பல்வேறு சொல்லமைப்புகள் காணப்படுகின்றன(எதில்). மேலும் அ,இ,உ என்னும் சுட்டெழுத்தை அடுத்து வரக்கூடிய அக் கூட்டத்தில், அவ் இடத்தில் போன்ற நிலைகளிலும் பிரித்தும் சேர்த்தும் எழுதப்படுகிறது இவற்றையெல்லாம் ஒரே ஒழுங்கிற்குக் கொண்டுவந்த பிறகே சொல்லடைவினை உருவாக்க வேண்டும்.

தமிழில் மொழியியல் விதிப்படி துணைவினைகள்(Auxiliary Verb), ஒட்டுகள்(Affixes) பிரித்து எழுதக்கூடாது என்ற நிலை இருக்கிறது. தான் என்பது இரண்டு நிலைகளில் வரும் எனவேதான் ஆகவேதான் என்னோடுதான் போன்ற உறுதிப்பொருளிலும் வரும், தான் என்று தன்னைக் கூறும்போதும் வரும். மேலும் பொருள் மயக்கம்(Ambiguity) வரக்கூடிய வேலை(வேல்+ஐ=வேலை,வேலை(Work)) அவரை, வருட, காலை, ஓட்டை, பாலை, வில்லை, சொத்தை, சூலை, காதை, கூட்டை இதுபோன்ற சொற்களையும் தெளிவுபடுத்த வேண்டும். (எதற்காக)

சொற்பட்டியலில் வேர்ச்சொல்லை அடிப்படையாக வைத்து உருவாகும் சொற்களை அடையாளம் காணமுடியும் அப்போது ஒரு வேர்ச்சொல்லில் இருந்து உருவாகும் சொற்களை ஒரே வருகையில்(Occurance) கொண்டுவரமுடியும். அவ்வாறு வரும்போது தமிழ் வேர்ச்சொல்லில் சில இடர்ப்பாடுகள் வருகின்றன. வந்தான், வருகிறான், வந்துகொண்டு, வராமல்... போன்று வரும்போது 'வ' குறில் எழுத்தில் வரிசைப்படுத்திக் காட்டும் ஆனால் இதன் வேர்ச்சொல் 'வா' என்பதாகும் இதேபோல் பல்வேறு வேர்ச்சொல் மூலவடிவ மாற்றமடையும் வினைகளும் மாற்றமடையாத வினைகளும் தமிழில் உள்ளன.

பொருள் மயக்கம்(Ambiguity) தரக்கூடிய சொற்கள் தமிழில் நிறைய உள்ளன. அவற்றுள் சில படி, காலை, போன்றனவாகும். ஒருசொல் பலபொருள், பலசொல் ஒருபொருள் என்னும் நிலையிலும் தமிழில் பொருண்மை நிலையில் சொல் பகுப்பு கண்டுணரப்படவேண்டும்.

தீர்வு

சொல்லடைவு உருவாக்கும்போது மனித உழைப்பால் முந்திருத்தம்(Pre-Editing) அல்லது பின்திருத்தம்(Post-Editing) செய்யவேண்டும். முன் திருத்தமே நலம். ஒரு படைப்பினை எடுத்துக்கொண்டால் அதில் உள்ள சொற்களின் கடைசி ஒற்றுகள் (க,ச,த,ப ஆகியவற்றால் முடியும் சொற்கள் மட்டும்) அனைத்தையும் நீக்க வேண்டும் அல்லது அதற்கு ஒரு விதி(Rule) அமைக்கவேண்டும். இதில் பொதுவான ஒரு தீர்வே கூறப்படுகிறது ஆனால் சில விதிவிலக்குகள் வரும்.

ஒரு வேர்ச்சொல்லுடன் சேரும் வேற்றுமைகள், சாரியைகள் போன்றவற்றைப் பிரித்து வரிசைப்படுத்த, கணினிக்கு உருபன் பகுப்பாய்வையும் சொல்லித்தரவேண்டியுள்ளது. தமிழில் உருபனியல் பகுப்பாய்விகள் உருவாக்கப்பட்டுள்ளன. அவற்றைப் பயன்படுத்திச் செயல்படுத்துவதால் இப்பிரச்சனையும் தீர்க்கப்படுகிறது.

கணினியில் Vb.net மொழிக்கருவியில் ரிச் டெக்ஸ்ட் ஆகத்தான் (Rich Text Box) படைப்புகளை உள்ளீடாகக் கொடுக்கமுடிகிறது. இதனால் படைப்பில் உள்ள பக்க எண்களைக் கொடுக்க முடிவதில்லை. அதற்கு Text

Control லட்சம் ரூபாய்க்கு மேல் விலை கொடுத்து வாங்கவேண்டிய சூழல் ஏற்படுகிறது. இதன் வாயிலாகப் படைப்பு எந்தெந்த பக்கங்களில் உள்ளதோ அதேபோல எந்த வரியெண்ணில் உள்ளதோ அதையும் அப்படியே பயன்படுத்த முடியும். இதனால் பக்க எண், வரி எண் ஆகியவற்றைக் கொடுப்பதில் உள்ள பிரச்சனை தீர்க்கப்பட்டது.

பயன்பாடு

- ❖ சொற்பட்டியல் தயாரித்தல் வாயிலாகத் தனிச்சொல் அதாவது தனிப்பெயர், வேற்றுமை ஏற்ற பெயர் தனி வினை, விசுவகளை ஏற்ற வினை என்னும் நிலைகளில் சொற்களை வகைபிரித்துச் செயல்படுத்துவதற்கு இது பெரிதும் துணைபுரியும்.
- ❖ சொற்களைத் தற்கால மொழியியல் அடிப்படையில் பெயர், வினை, அடை, ஒட்டு என்பன போன்ற கூறுகளில் வகைப்படுத்தி ஆராய முடியும். மேலும் ஒரு வேர்ச்சொல்லை அடிப்படையாக்கொண்டு எவ்வாறெல்லாம் சொற்களை உருவாக்கமுடியும் என்றும் வரிசைப்படுத்த முடியும்.
- ❖ வேர்ச்சொற்களுக்கானப் பட்டியலை உருவாக்க முடியும். ஒரே வேர்ச்சொல்லை அடிப்படையாகக்கொண்டு எத்தனை சொற்களையும், சொற்றொடர்களையும் உருவாக்கமுடியும் என்று கணக்கிட்டு ஆராய முடியும். சொல் நிலையிலும் தொடர் நிலையிலும் பொருண்மை நிலையிலும் பனுவல்களை ஆராய முடியும்.
- ❖ சொல்லின் ஒரு பகுதியைத் தேடுவதன் வாயிலாக ஒரு விசுவ எந்தெந்த சொற்களோடெல்லாம் சேரும் என்பதையும் கண்டறிந்து வரிசைப்படுத்த முடியும். இந்தச் சொல்லடைவு தமிழ் கார்பஸ் (CORPUS) தயாரிப்பதற்குப் பெரிதும் பயன்படும். சொற்பட்டியல், சொல்லடைவு ஆகியவற்றைக் கணினிவழி உருவாக்கும் முறைகளைப் பின்பற்றி பனுவல்களிலிருந்து அகராதிகளை உருவாக்கலாம்.
- ❖ பனுவலுக்கான பொருண்மையை இதன் வாயிலாக எளிமையாகக் கண்டறியலாம். மேலும் பல்பொருள் குறித்த ஒரு சொல், ஒருசொல் குறித்த பல்பொருள் குறித்து ஆய்வு செய்யவும் வகைப்படுத்தவும் இதனால் சாத்தியமாகின்றது. இச்சொல்லடைவுகளைக் கொண்டு இலக்கியச் சொற்கள் காலத்திற்கேற்ற சொற்கள், பிறமொழிச்சொற்கள், வட்டார வழக்குச் சொற்கள் எனப் பகுத்தறியப்படுகின்றன. சொற்களின் தேர்வு, பயன்படுத்தும் முறை, புதிய சொற்களை ஆக்கும் திறன், சொற்களுக்குப் புதிய பொருள் அளித்தல் போன்றவைகளையும் ஆராய்வதற்கு இது பெரிதும் உதவும்.
- ❖ கவிதை, கட்டுரை, சிறுகதை, நாவல் போன்ற எந்த ஒரு படைப்பையும் உள்ளீடாகக் கொடுத்துச் சொற்களைத் தனித்தனியாகப் பிரித்து ஒவ்வொரு சொல்லும் எத்தனை முறை பயின்றுவந்துள்ளது என்பதையும் அந்தச் சொற்களை அகரவரிசையாகவும், நிகழ்வெண்ணிக்கை அடிப்படையாகவும் வரிசைப்படுத்தவும் செய்யமுடியும்.
- ❖ வினா, உணர்ச்சி, கூற்று போன்ற வாக்கிய வகைப்படி வரிசைப்படுத்தவும் இலக்கியச் சொற்கள், பிறமொழிச்சொற்கள், தற்காலச் சொற்கள் எனப் பல்வேறு வகைப்பாடுகளில் வகைப்படுத்த முடியும். சொற்கள் புழக்கமும் அந்தச் சொற்கள் பயிற்றிடங்கள் அதாவது இந்தச் சொல் இந்தந்த இடங்களில்தான் வரும் என்று தற்கால இலக்கண விதிகளை உருவாக்குவதற்குப் பெரிதும் பயன்படும். ஒரு வேர்ச்சொல்லை அடிப்படையாக வைத்து உருவாக்கக்கூடிய சொற்களை வகைப்படுத்தும் Lemma Extractor உருவாக்கப் பயன்படும். பொருள் மயக்கம் வரக்கூடிய சொற்களுக்கு உடனடியாக எந்தப் பொருளில் பயன்படுத்தப்பட்டுள்ளது என்பதை உடனடியாக அறிந்துகொள்ளவும் வகைப்படுத்தவும் தொடரடைவு மிகவும் இன்றியமையாததாகும்.

முடிவாக

- ❖ கணினிவழிச் சொல்லடைவு மென்பொருள் உருவாக்குவதை இணையவழியாக எந்த ஒரு படைப்பையும் இம்மென்பொருளைப் பயன்படுத்திச் சொல்லடைவினையும் தொடரடைவினையும் உருவாக்கிக்கொள்ள வழிவகைச் செய்யப்பட வேண்டும். தமிழில் தரவுத்தளங்கள்(Database) பல்வேறு நிலைகளில் அமைக்கப்பட்டு அனைவரும் பயன்படுத்தும் வகையில் இலவசமாக இணையத்தில் அளிக்கப்படவேண்டும். தமிழில் தற்போது தேவைக்கேற்ப மின் அகராதிகளும் மின்சொற்களஞ்சியங்களும் உருவாக்கப்பட்டுவருகின்றன.
- ❖ சொல்லடைவு மென்மம் உருவாக்கி செயல்படுத்தும்போது அதனைச் சங்க இலக்கியத்திற்கென்றும் தற்காலப் படைப்புகளுக்கென்றும் இரண்டு நிலையில் அமைக்க வேண்டும். தனித்தனியாக ஒவ்வொரு படைப்புக்கும் சொல்லடைவு என இல்லாமல் எவ்வகைப் படைப்பையும் கணினிவழிச் சொல்லடைவினை உருவாக்கும் மென்பொருளை உருவாக்க வழிவகை செய்யப்படவேண்டும்.
- ❖ கணினிவழி எந்த ஒரு படைப்பையும் சொல்லடைவு(Index), தொடரடைவு(Concordance) உருவாக்குவதற்கு முழுமையாகச் செயல்படும் மென்மங்கள் வெளிவரவில்லை எனலாம். இவ்வாறு வெளிவரவேண்டும் அதற்கான முயற்சிகளை மேற்கொள்ளவேண்டும் என்பதே இக் கட்டுரையின் நோக்கம். சொல்லடைவு, தொடரடைவு, பொருளடைவு உருவாக்குகின்ற நிலையில் இணையவழி நூலடைவு, இணையவழி தமிழியல் ஆய்வுகள் அடைவு போன்றவை உருவாக்கப்படவேண்டும். தேவை ஏற்படின் இக் கட்டுரையோடு கட்டுரையாளர் உருவாக்கிய சொல்லடைவுக் கருவியின் செயல்பாட்டைக் கணினியில் செயல்படுத்திக்காட்டப்படும்.
- ❖ கணினிவழித் தொடரடைவில் சொற்கள், சொல் விளக்கங்கள், சொல் குறுக்கீட்டு நோக்குகள் ஆகியவையும் இடம்பெறும். தொடரடைவியினை உருவாக்குதலின் வாயிலாகப் பல நன்மைகள் ஏற்படும். ஒரே சொல் பல்வேறு தொடர்களில் அமையும் பல்வேறு அமைப்பினைக் கண்டறிய முடியும். தலைப்புச்(HeadWord) சொற்களைப் பகுப்பாய்வு செய்யவும், சொற்களின் வருகை எண்ணிக்கையை ஆய்வு செய்யவும், மரபுத் தொடர்களைக் கண்டறிவதுடன் அவற்றைப் பகுப்பாய்வு செய்யவும் தொடரடைவு வழி வகை செய்யும். அத்துடன் தொடரடைவுகளிலிருந்து சொல்லடைவுகளையும் சொல் பட்டியல்களையும் பெற முடியும். கணினிவழிச் செய்யப்படும் இத்தொடரடைவுப் பணி எந்திர மொழி பெயர்ப்புக்கும் பயன்படும். எனவே கணினி வழித் தொடரடைவு என்பது பல்வேறு நிலைகளில் மொழி ஆய்விற்குப் பயன்படும் கருவி.

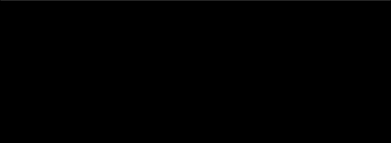
7

இணையத்தொழில்நுட்பத்தில்

தமிழ் மொழி

மற்றும்

திறவூற்று செயலிகள்



Fostering Tamil Web Communities for Mining Tamil Web Pages

Dr. A. Vijaya Kathiravan, R. Vidhya
Dept. of Master of Computer Applications,
KSR College of Technology, Tiruchengode - 637 215
Contact: vijayakathiravan@hotmail.com,
vijayakathiravan@gmail.com, Call: +91-9244217777

Abstract

“There are three popular qualities identified with Tamil people in 20th century: Creativity, intelligence and artistic thinking“. With the growing Tamil interest and Internet, the amount of Tamil data doubles every 12-14 months and will increase even more dramatically in the coming year. With an enormous amount of Tamil data stored in web databases and warehouses, it is increasingly important to develop powerful tools for analysis of such Tamil data and mining interesting patterns from it. There is a strong interest in employing methods of data mining to generate models of Tamil related web pages forming web communities. The main intention of this paper is to establish cyber community mining technique for Tamil domain and to identify the Tamil resources available in the free web. This paper proposes a new initiative for forming Tamil web communities with concise introduction about community mining. It also groups research publications and literatures in Tamil using bibliometric analysis. This community mining will yield benefits to all Tamil lovers, who want to be well-versed in a Tamil domain of his own interest. By forming people communities (i.e., people belonging to similar interest) using social network analysis, the domain knowledge in Tamil can be shared. Hence, web community mining may play an important role in forming Tamil Web Communities for gathering Tamil resources and documents of similar interest.

Keywords: Web communities, Tamil communities, social network analysis, community mining, web mining, web structure mining, citation, co-citation and bibliometric coupling.

1. Introduction

The ability to train computers to predict properties based upon knowledge of Tamil computing for Tamil data offers the prospect of automatically screening massive libraries of other language information to produce prediction. Tamil computing is an application of information technology in Tamil to the storage, management and analysis of Tamil information. Tamil information consists of Tamil databases, Tamil articles, Tamil publications, Tamil literatures, Tamil news, agricultural, health care, scientific information and other Tamil related information. With the increase in Tamil interest in web, computer-based Tamil databases, operating systems, hardware, software, tools, search engines and mining techniques have become very essential. Nowadays, web mining outperforms a significant position in information retrieval process using search engines, in which the main ingredients are web content mining, web structure mining and web usage mining. About 40% of the people are likely to surf the web using hyperlinks. Community mining is transpired from web structure mining that aims to

utilize the hyperlinked structural pattern available in web provided that it has some meaningful information hidden inside. A web community refers to group of web pages sharing a common interest like Tamil implicitly or explicitly. Using people communities, people in similar profession can group together as virtual teams to accomplish their tasks.

2. Tamil Community Mining

A Tamil web community is a set of Tamil web pages that provide resources on a specific topic in Tamil. By modeling the Web as a graph and performing several operations on it, it was able to separate the Web in sets of related items in which Tamil community is extracted. The different types of association relationship existing in a community are citation, co-citation and bibliographic coupling. Those have been described in fig 1.

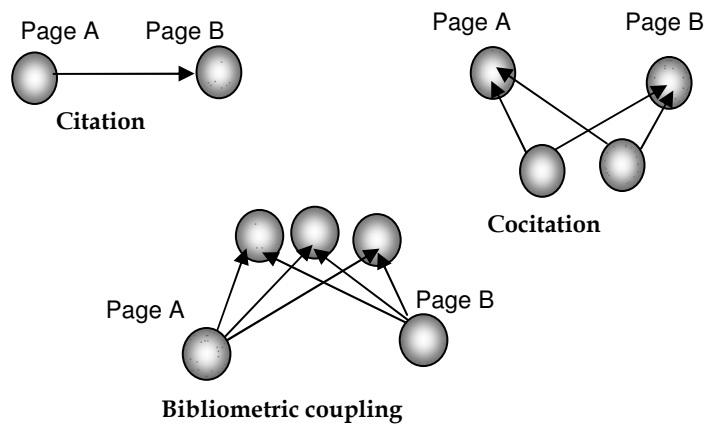


Fig 1. Types of associations in Community

Using the above associations, bibliometric coupling finds patterns in citation graphs; sociometric finds patterns in social networks; collaborative filtering finds patterns in rank graphs; webometric finds patterns in web page links. The main scope of community mining is to measure the similarity of web pages on the web graph and to extract the meaningful communities through the link structure pattern.

2.1 DEFINITIONS OF COMMUNITY

Several different definitions of community were also raised in the literature that has been explained in

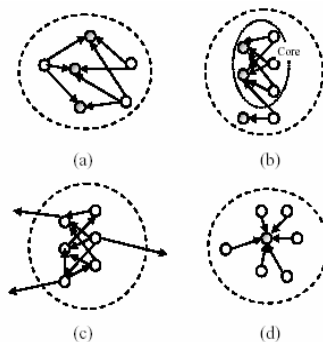


Fig 2. Several Different Definitions of Community

- (a) A web community is a number of representative authority web pages linked by important hub pages that share a common topic.
- (b) A web community is a highly linked bipartite sub-graph and has at least one core containing complete bipartite sub graph.
- (c) A set of web pages that linked more pages in the community than those outside of the community could be defined as a web community.
- (d) A research community could be based on a single most cited paper and contain all papers that cite it. While each of the above definition characterizes some essential properties of a community, it makes the community mining task rather difficult because of a lack of uniform definition.

3. Discovering Communities

In this paper, a community on the web is defined as a cluster of web pages, which share common topics. However, there are many ways to detect the clusters of web communities. One of the key distinguishing features of the algorithms has to do with the degree of locality used for assessing whether or not a page should be considered a community member. On the one extreme are purely local methods, which consider only the properties of the local neighborhood around two vertices to decide if the two are in the same community. Global methods operate at the other extreme, and essentially demand that every edge in a Web graph be considered in order to decide if two vertices are members of the same community. Broder et al. [1] reported on an algorithm of clustering web pages based on the contents. This approach can be applied not only to hypertext but also plaintext. However, indexing web pages accurately is difficult because the contents of web pages are not always meaningful. In contrast to the content-based approach, links in web pages can be reliable information because they reflect human judgment. Botafogo and Shneiderman [2] proposed an idea for abstraction called aggregate based on graph theory.

Their algorithm removes 'indics' (nodes with high number of out-links) and 'references' (nodes with high number of in-links) iteratively to clear the graph. However, removed nodes often become very important elements to understand the web. On the other hand, Kumar et.al., [3] defined a community on the web as a dense directed bipartite subgraph, and discovered over 100,000 communities. However, the scale of subgraphs depends on its parameters. This implies the difficulty in detecting communities from the web, since the communities are often somewhat related with each other. As another use of links, Kleinberg [4] and Brin and Page [5] used the link structures for ranking web pages. Their main idea was based on mutually reinforcing that the more a web page is referred, the more authoritative the web page becomes, and the higher the web page ranks. The highly ranked web pages tend to be the representative web pages of communities. There are several data structures such as biconnected components, strongly connected components, bipartite graph (BG), dense bipartite graph (DBG) and Complete bipartite graph or bipartite core (CBG) employed for forming communities.

4. Terminologies In Web Community

Bipartite graph (BG). A bipartite graph $BG(T, I)$ is a graph, whose node-set can be partitioned into two non-empty sets T and I . Every directed edge of BG joins a node in T to a node in I .

Dense bipartite graph (DBG). Let p and q be nonzero integer variables and t_c and i_c be the number of nodes in T and I , respectively. A $DBG(T, I, p, q)$ is a $BG(T, I)$, where (i) each node of T establishes an edge

with at least p ($1 \leq p \leq ic$) nodes of T , and (ii) at least q ($1 \leq q \leq tc$) nodes of T establish an edge with each node of I .

Complete bipartite graph (CBG). A CBG (T, I, p, q) is a DBG (T, I, p, q) , where $p = ic$ and $q = tc$.

K L M K L M

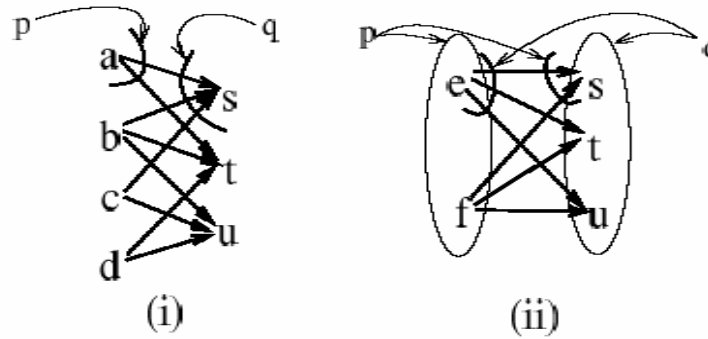


Fig 3. (i) DBG (T, I, p, q) (ii) CBG (p, q)

Fig 3 shows the difference between a DBG (T, I, p, q) and a CBG (p, q) .

Community hierarchy

Let the variable `num_levels` denote the number of levels in a hierarchy for a given data set. A community is denoted with $C(i, j)$, where i ($1 \leq i \leq \text{num_levels}$) is a nonzero integer value that denotes the level and j is an integer value which denotes unique community identifier at level i . Then, If $i = 1$, members of $C(i, j)$ are the web pages. If $i > 1$, members of $C(i, j)$ are the communities at level “ $i-1$ ”.

Community $(C(i, j))$ Let p_i and q_i be integer variables that represent threshold values. The community $C(i, j) = T$, if there exist a DBG (T, I, p, q) over a set of nodes at level “ $i-1$ ” with $p \geq p_i$ and $q \geq q_i$.

Cocite Let p_i and p_j be pages. $\text{Cocite}(p_i, p_j) = \text{true}$, if $|\text{child}(p_i) \cap \text{child}(p_j)| \geq \text{cocite_factor}$, where `cocite_factor` represents a nonzero integer value.

Relax cocite Let T be the set of pages and P_j be the another page ($P_j \in T$). For any page $P_i \in T$, if $\text{relax_cocite}(T_i, T_j) = \text{true}$ if $|\text{child}(P_j) \cap \text{child}(T)| \geq \text{relax_cocite_factor}$. Here, `relax_cocite_factor` is nonzero integer variable and $\text{child}(T)$ contains the children of the pages of T .

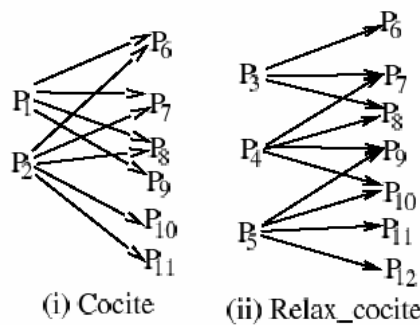


Fig 4. Depiction of Cocite and Relax_cocite

Max flow min cut theorem. According to Floyd and Fulkerson’s Max flow min cut Theorem, an ideal community $C \subseteq V$ can be identified by calculating the s-t minimum cut using appropriately chosen source and sink nodes. In recurse over algorithm, community obtained in one iteration is used as input to next iteration. Finally, mirroring pages in the community are eliminated using “shingling” method. Quite a lot of work has been done in mining the implicit communities of users, web pages or scientific literature from the Web or document citation database using content or link analysis. Forming Communities among users is called as people community. Using bibliographic references in scientific literatures and document citations, scientific literature community can also be formed, which are described in detail.

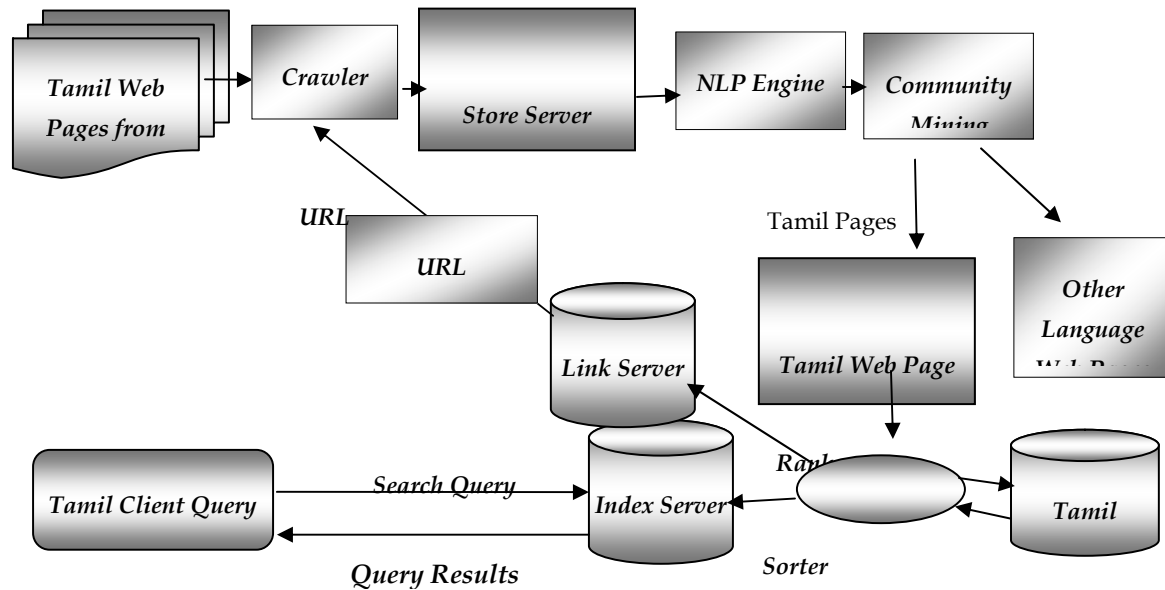


Fig. 5: Architecture of Search Engine for mining Tamil Web Communities

5. Proposed Algorithm for Evaluating Tamil Communities

5.1 Main phases

Phase1. Collect web pages related to a Tamil domain.

Phase2. Discover communities on the web for all domains.

Phase3. Discover established relations among the communities.

Phase4. Discover future enhancements among the communities.

Phase5. Visualize inter and intra relationship of communities.

5.2 The Detailed Process

Phase1. Preparations: First of all, let a user decide a Tamil domain, which she/he want to explore thoroughly.

Then, source web pages D are collected by using any conventional search engine like Google. Here, web pages of Google's output for the query are downloaded.

Phase2. Discover Communities: Use community trawling algorithm or s-t max flow min cut algorithm for community mining. For surveying the picture of communities with future enhancements among communities, only centered web pages in communities are used instead of all the web pages. The centered web page named as core-page is extracted as follows.

1. Count the frequency of links included in D .
2. Regard the top $N1$ links C as the 'core-pages' of communities.

Phase3. Discover Established Relations: Measure the relations among core pages by counting the number of co-citations, and regard strong relations as established links. The process is as follows.

1. For every pair of two core-pages in C , count the number of links included in both the core-pages.
2. Regard the top $N2$ pairs as established links $L1$ (solid lines in Fig. 6).

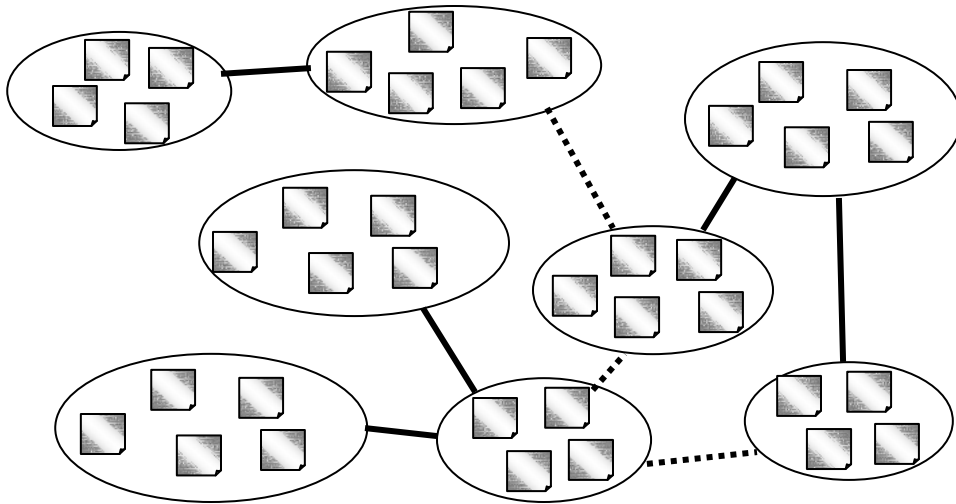


Fig 6. Visualization of established and future links in Tamil communities

Phase4. Discover Future Enhancements: Measure the relations among core-pages by counting the number of co-citations, and regard weak relations as future links. The process is as follows.

1. For every pair of two cores in C except for $L1$, count the number of links included in both the cores.
2. Regard the top $N3$ pairs as future links $L2$ (dotted lines in Fig. 6).

The movement of communities is shown by established and future links. Therefore, future enhancements are expressed by the combination of these two kinds of relations.

Phase5. Visualization: Core-pages and its relations (C , $L1$, and $L2$) are visualized into 2-dimensional interface to piece out the connections of communities and to understand the potential needs or demands.

Conclusion and Future Work

In this paper, the importance of discovering communities in Tamil for gathering resources has been strongly insisted. Using visualization, an idea for discovering established web pages in a Tamil community is mined by chaining primitive communities to understand potential needs or demands. It could find communities of adaptive granularity. By using techniques such as community clustering, social network analysis and bibliometric analysis, the Tamil world can get more benefits in future. For efficient community mining, algorithms based on machine learning may be used in future. Since this is a general idea that could potentially be used in any application scenarios, where the data can be abstracted to a graph structure, this may be applied to any domain. Its usefulness in other environments such as human relationship network, newsgroups, communication network, etc can be tested and the evolution of communities can be analyzed in future. Since the objects and links in a network are usually dynamic, changes of communities in a time series manner can be observed. From this idea of implementing community mining in Tamil, it has been confirmed that communities are the useful means to share resources and knowledge among people from different community domains.

People community and scientific literature community may act as a powerful knowledge acquisition tool for the Tamil community. The ideas and methods presented in this study can be proved useful by analyzing more domain communities. Future research in community mining focuses on generalizing the notion of community parameterized with coupling factor low for weakly connected communities, high for highly connected communities and optimal for ideal community. Co-boosting is a method for information retrieval from an unlabelled data. By combining this technique with community mining in Tamil, variety of Tamil based community like social community, literature community, political community, student community, Tamil cinema and songs community, research community and other Tamil related communities may be formed in future.

Acknowledgement

"Thou art the real goal of human life, we are yet but slaves of wishes putting bar to our advancement. Thou art the only God and Power to bring us upto that stage". I sincerely thank INFITT-2010 Conference Selection Committee for having given us this opportunity. My extended gratitude goes to our college superiors, colleagues, friends and family members.

REFERENCES

1. Broder, A. Z., Glassman, S. C., Manasse, M. S., "Syntactic Clustering of the Web", Proc. World Wide Web Conference, 1997.
2. Botafogo, R. A., Shneiderman, B, "Identifying Aggregates in Hypertext Structures", Proc. ACM Conference on Hypertext, p.63-74, 1991.
3. Kumar. R., Raghavan, P. Rajagopalan. S., Tomkins, A., "Trawling the Web for Emerging Cyber-Communities", Proc. World Wide Web Conference, 1999.
4. Kleinberg, J. M., "Authoritative Sources in a Hyperlinked Environment", Proc. ACM- SIAM Symposium on Discrete Algorithm, p. 668-677, 1998.
5. Brin, S., Page, L., "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Proc. World Wide Web Conference, 1998.

6. Furen Lin, ChunHung Chen, KuoLung Tsai, "Discovering Group Interaction Patterns in a Teachers Professional Community", Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), IEEE, 2002.
7. Wen-Jun Zhou, Ji-Rong Wen, Wei-Ying Ma, Hong-Jiang Zhang, "A Concentric-Circle Model for Community Mining in Graph Structures", Technical Report in Microsoft Research, MSR-TR-2002-123, Nov. 15, 2002.
8. P. Krishna Reddy and Masaru Kitsuregawa, "An approach to relate the web communities through bipartite graphs", Institute of Industrial Science, The University of Tokyo, Japan, 2001.
9. Naohiro, Matsumura1, Yukio Ohsawa, Mitsuru Ishizuka, "Future Directions of Communities on the Web", School of Engineering, University of Tokyo, Japan, 2000.
10. Alexandrin Popescul, Gary William Flake, Steve Lawrence, Lyle H. Ungar, C. Lee Giles, "Clustering and Identifying Temporal Trends in Document Databases", in IEEE Advances in Digital Libraries, ADL 2000.
11. Dmitry Zelenko, Chinatsu Aone, 2006, *Discriminative methods for Transliteration*, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing(EMNLP2006), pages612-617, 2006.
12. Surya Ganesh, Sree Harsha, Prasad Pingali, Vasudeva Varma, *Statistical transliteration for CrossLangauge Information Retrieval using HMM alignment and CRF*, The Second International Workshop on Cross Lingual Information Access-Addressing the Informaion Need of Multilingual Socoeties, 2008.
13. Sathiya Keerthi S, Sundararajan S, *CRF versus SVM-Struct for Sequence Labeling*, Yahoo Research technical report, 2007.
14. Taskar B, Lacoste-Julien S, and Klein D., *A Discriminative Matching Approach to Word*, 2005.



Dr. A. Vijaya Kathiravan is an Asst. Professor in Department of Computer Applications, K.S.R. College of Technology, Tiruchengode, TN, INDIA. She received her M.Phil. in Computer Science from Bharathiar University, Coimbatore and recently she awarded her doctoral degree in University of Madras, Chennai. Her research interests include information retrieval, web communities, machine learning, data mining, data structures, text mining, NLP, social network mining, leadership assessment and human resource management.

பன்னாட்டு இணைய முகவரி முறையில் தமிழ்

நெடுங்கணக்கைப் பயன்படுத்துவதில் உள்ள

சவால்கள் (IDN- தமிழ் முகவரிகள்)

டான் டின் வீ

சிங்கப்பூர் தேசிய பல்கலைக்கழகம்

எஸ் சுப்பையா, எஸ் மணியம்

i-DNS.net International Inc.

பாலோ ஆல்டோ, கலிஃபோர்னியா, அமெரிக்க ஐக்கிய நாடுகள்

மின்னஞ்சல்: maniam@i-dns.net

பிழிவு

இணையத்தில் இணைந்துள்ள கணிப்பொறிகள் அல்லது சாதனங்களின் இணைய முறைமை முகவரி களை (IP Address) அர்த்தமுள்ளதாகவும் எளிதில் நினைவில் வைத்துக்கொள்ளக்கூடியதாகவும் உள்ள பெயர்களோடு பொருத்தும் விவரங்களைக் கொண்ட ஓர் அனைத்தளாவிய தரவுத்தரமே இணையத்தின் இணைய முகவரிகள் பெயர் அமைப்பு ஆகும். ஆனால், இதுவரை இந்த அமைப்பில் ASCII தரநிலை யிலுள்ள எழுத்துக்களும் எண்களும் நிறுத்தற்குறிகளும் மட்டுமே ஏற்கப்பட்டன. பெரும்பாலும் ஆங்கிலத்தை இணைப்பு மொழியாக்கக் கொண்ட கல்விச் சமூகத்திடம் மட்டுமே இணையம் முடங்கிக் கிடந்தவரை இது சரியானதாக இருந்திருக்கலாம்.

ஆனால் 1990களின் நடுவிலிருந்து, இணையம் பன்மொழித்தன்மை பெறத் தொடங்கிவிட்டது. எல்லா நாடுகளிலும் உள்ள எல்லா மொழி பேசுவோர்க்குமான உள்ளடக்கம் இணையத்தில் வரத்தொடங்கின. வலை உலாவிகளும் அதற்கேற்ப எல்லா மொழிகளின் எழுத்துக்களையும் காட்சிப்படுத்தும் திறன்களோடு வெளியாகத்தொடங்கின.

1998 இலிருந்து இணையத் திணைப் பெயர் அமைப்புகளை பன்னாட்டுமயமாக்குவதற்கான நடவடிக்கைகள் தொடங்கின. ஏற்கனவே உள்ள லத்தீன் எழுத்து முறையிலான திணைப் பெயர் அமைப்புகளோடு இசைந்து செல்லும் வகையில், பன்னாட்டு திணைப் பெயர் அமைப்புகளை (IDNs) உருவாக்கும் முயற்சி எடுக்கப்பட்டது. உருவாகி வந்த IDN கள், அமெரிக்காவின் ஆதிக்கத்தின் கீழ் இருந்தது. 1998 இல் தொடங்கப்பட்ட Internet Consortium for Assigned Names and Numbers (ICANN) அமைப்பு இந்த புதிய முயற்சிக்கான தரநிலைகளையும் செயல்முறைகளும் உருவாக்கி அங்கீகாரமும் அளிப்பதற்கு கிட்டத்தட்ட பத்தாண்டுகளை எடுத்துக்கொண்டது. முதலில் பகுதி IDNகள் அனுமதிக்கப் பட்டு, இரண்டாம் அல்லது மூன்றாம் நிலை திணைகளில் ஏற்கப்பட்டன. .COM போன்ற உயர்நிலை ASCII திணைப் பெயர்களில் இவை அனுமதிக்கப்பட்டன. ஆனால் இந்த அரைகுறை பன்னாட்டு மயமாக்கம் பன்மொழி சமூகத்தில் ஏற்கப்படவில்லை. முழுமையான IDNகளை அறிமுகப்படுத்துவதை எதிர்த்து சிலர் செய்த எதிர் முயற்சிகளின் காரணமாக, முழு IDN களுக்கு உத்தரவாதம் அளித்திருந்த போதிலும், ICANN ஆல் அதை அளிக்கமுடியவில்லை.

இப்போது சில திணைப்பெயர்கள் மட்டும் அறிவிக்கப்பட்டுள்ளன. இந்த வரலாற்று முக்கியத்துவம் வாய்ந்த பெயர் இடுதல் முறையில் தமிழர்கள் தங்களது பங்கினை நிச்சயமாக எடுத்துக்கொள்ள வேண்டும். IDN பணிக்குழுக்கள் என்றால், நாம் மிகவும் தொடக்க காலத்திலேயே உத்தமத்திலேயே

செய்யத்தொடங்கிவிட்டோம். இப்போது, இணைய உலகில் தமிழ் தன் சொந்த திணைப்பெயர்களோடு விளங்குவதற்காக, 12 ஆண்டுகாலமாக நாம் எடுத்துவந்த முயற்சிகளைப் பற்றி விரிவாகப் பார்ப்போம்.

1. தற்போதைய DNS அமைப்பில் பன்மொழி வரம்புகள்

இணைய முகவரிகளை சுலபமாக நினைவில் வைத்துக்கொள்ளும் பொருட்டு இணைய திணைப் பெயர் அமைப்பு (DNS) உருவாக்கப்பட்டிருக்கிறது. எடுத்துக்காட்டாக, www.yahoo.com என்பது போன்ற எளிதில் நினைவில் வைத்துக்கொள்ளக்கூடிய எழுத்துக்களும் எண்களும் மட்டும் உள்ள பெயர்களை, 137.132.19.1 என்பது போன்ற (இது IP பதிப்பு 4 ஐச் சேர்ந்தது) வெறும் எண் தொடர்களுக்கு பொருத்தி, அவற்றுக்கு பதிலியாக பயன்படுதலாம். துரதிருஷ்டவசமாக, அனைத்தளாவிய தன்மையை முதன்மை யாக கருதி, (LDH (letters-digits-and-hyphen) என்று அழைக்கப்படும் ASCII எழுத்துக்கள் மற்றும் எண்கள், ஹைபன் ஆகியவற்றை மட்டுமே திணைப் பெயர்களுக்கு பயன்படுத்துவது என்று முடிவு செய்யப்பட்டிருந்தது. இவ்வாறாக அர்த்தம் புரியாத எண் தொடர்களுக்கு பதிலாக, அர்த்தம் பொதிந்த எழுத்து-எண் தொடர்களை பயன்படுத்தியதால், திணைப் பெயர் முறையை இணைய வழங்கிகளுக்கும் மின்னஞ்சல் முகவரிகளுக்கும் வலையக முகவரிகளுக்கும் பயன்படுத்தத் தொடங்கி, உலகம் முழுவதும் பிரபலமாகிவிட்டது. ஆங்கிலம் பேசாத நாடுகளில் இணையத்தையே அணுகமுடியாத தடைக்கல்லாக மாறிவிட்டது.

2. இணைய முகவரி அமைப்பை பன்னாட்டுமயமாக்கல்

இணையம் பன்மொழித்தன்மையைப் பெறத்தொடங்கிய பிறகு, இயற்கையாகவே அதன் திணைப் பெயர்களையும் பன்மொழியில் அளிப்பது தொடர்பான முயற்சி 1998 இல் தொடங்கியது. ஆனால் DNS ஐ அடுத்த நிலைக்குக் கொண்டுசெல்லும் முயற்சி கடுமையானது. DNS முறைமை என்பது வேறு பல இணைய முறைமைகளின் மீது பாதிப்பு ஏற்படுத்தக்கூடியது. எடுத்துக்காட்டாக, வலையகங்களின் முகவரி, மின்னஞ்சல் முகவரி போன்றவை அது பாதிக்கிறது. எனவே வலை உலாவிகளும் மின்னஞ்சல் மென்பொருள்களும் மாற்றம் கொள்ளவேண்டியவரும். DNS சர்வர்களில் உள்ள சர்வர் மென்பொருள்கள் உலக இணையத்தை பின்னிப்பிணைத்து வைத்திருக்கிற நிலையில், அவற்றில் மாற்றம் செய்யவேண்டும் என்பது மிகவும் செலவுபுடிக்கக்கூடிய, படிப்படியாக, பரவலாக மேற்கொள்ளப்படக்கூடிய செயல் பாடாக மாறிவிடுகிறது.

1998 இல், சிங்கப்பூர் தேசிய பல்கலைக்கழகத்தில், இன்டர்நெட் ரிசர்ச் அண்டு டெவலப்மென்ட் யூனிட்டில், இந்தக் கட்டுரையின் முதல் ஆசிரியர் டான் டின் வீ, இந்த பிரச்சினைக்கு தீர்வு அளிக்க முன்வந்தார். முடிந்தவரை முன்னிலை-ஏற்பு நிலைக்கு உட்பட்ட (backward compatible), எதிர்காலத்தில் நீட்டிக்கப்படக்கூடிய, மென்பொருள்களில் அதிகம் மாற்றம் செய்யத்தேவை இராத ஒரு DNS முறை குறித்து தீர்வு உருவாக்கி, செயற்படுத்திக்காட்டினார்.

வலுவான பன்மொழி திணைப் பெயர் அமைப்பை - iDNS - உருவாக்கி, பன்னாட்டு சோதனைக் களத்தில் வைத்து பரிசோதிக்க, ஆசிரியர் பசிபிக் உருவாக்க குழு ஒன்று உருவாக்கப்பட்டது. 1998 ஆகஸ்ட் முதல் டிசம்பர் வரையிலான காலத்தில், தலைவர் ஹாங்காங், இந்தியா, கொரியா, சீனா போன்ற நாடுகளுக்குச் சென்று, சோதனைநிலை அமைப்பு முறையை செயல்விளக்கம் அளித்துக் காட்டினார். இந்தத் திட்டத்துக்கான பன்னாட்டு ஆதரவைத் திரட்டத்தொடங்கினார்.

1999 இறுதியில், இது தொடக்கநிலை தொழில்நுட்பங்களை உருவாக்கிய, பல்கலைக்கழகத்திலிருந்து உருவான BIX Pte Ltd நிறுவனம் i-DNS.net International Inc. என்கிற தனியார் நிறுவனமாக வடிவம் பெற்றது. இதற்கிடையில், IETF இல், பன்னாட்டுமயமாக்கப்பட்ட திணைப் பெயர் IDN பணிக்குழு உருவாக்கப்பட்டது, iDNS க்கான அடிப்படையான தேவைப்பாடுகளை இனமறியும் செயல்பாடுகள் தொடங்கின. இந்தப் பணிக்குழு ஒரு தரநிலை பின்தொடர்வு பணக்குழுவாக மாறி, பன்னாட்டு செயல்பாடுகள் விரைவுபடுத்தப்பட்டன. இறுதியில், இணையத்தின் தரநிலைகளுக்கான அங்கீகாரத்

தலைமைத்துவமான இன்டர்நெட் எஞ்ஜினியரிங் டாஸ்க் ஃபோர்ஸ் (IETF) இன் INDA தரநிலைகளை உருவாக்குவதில் அது முடிந்தது. 2003 இலேயே IDNA தரநிலைகள் உருவாக்கப்பட்டும் கூட, அமெரிக்க அரசின் ஆணைக்கிணைங்க இணைய திணைப்பெயர்களையும் எண்களையும் ஒதுக்கீடு செய்யும் அமைப்பான இன்டர்நேஷனல் கார்ப்பரேஷன் ஃபார் அசைண்ட் நேம்ஸ் அண் நம்பர்ஸ் (ICANN) இதற்கு ஏற்பு அளிக்கவும் இதை அமல் படுத்தவும் கூடுதல் காலத்தை எடுத்துக்கொண்டது.

3. இதற்கு இவ்வளவு காலம் பிடித்தது ஏன்?

IDN ஐ நிறுவுவதற்கான ஐந்து ஆண்டு கால உழைப்புகள், தொழில்நுட்ப ரீதியிலும் வர்த்தக ரீதியிலும், எண்ணற்ற மாறுதல்களுக்கு உள்ளாகிக் கொண்டே இருந்தது. நிறைய நிறுவனங்கள் இதில் இறங்கின, ஆனால் அதில் பல காணாமல் போயின. 2000களின் தொடக்கத்தில், இந்த விவகாரம் சூடுபிடித்திருந்த நிலையில், குறைந்த பட்சம் ஒரு டஜன் தீர்வுகளாவது இதற்காக சந்தையில் நிலவின.

2010 இல், இன்று, IDN இயக்கம் சில சிக்கல்களைச் சந்தித்த பிறகு, ICANN அமைப்பு இறுதியாக, IERE தரநிலைகளை (RFCகள் 3490, 3491 மற்றும் 3492) அடிப்படையாக்கக் கொண்டு ஒரு IDN திட்டத்தை உருவாக்கியது. இதன் மூலமாக, தொடக்கம் முதலே IDN ஐ ஆதரித்து உழைத்துவந்த முன்னோடிகளை அது புறக்கணிக்கும் வரலாற்றுப்பிழையைச் செய்தது. உண்மையில், 2000களில் (RFC2825 , தேதி மே 2000 பார்க்க) IDN களின் தேவை குறித்து அது அங்கீகாரம் செய்வதில் காட்டிய தாமதமும் அலட்சியமும் குறிப்பிடத்தக்கது. இணையத்தின் தனித்தன்மைக்கு பாதகமாக போகலாம் என ICANN உம் IABஉம் அதைப் பார்த்தன. (பார்க்க RFC2826 , தேதி 2000). அதுமட்டுமல்லாமல், அமெரிக்காவுக்கு வெளியே யிருந்து வரும் முயற்சிகளை வெற்றிகரமாக தடுத்து நிறுத்தி, தனது மேலாதிக்கத்தை பாதுகாக்க முனைந்தது. இதன் காரணமாகவே, பத்தாண்டுகள் இந்த தாமதம் விளைந்தது.

4. பன்னாட்டு இணைய பெயர்களில் தமிழைப் பயன்படுத்துவது தொடர்பான சவால்கள்

IDN களைப் பயன்படுத்துவதில் பல மொழிகளிலும் பல சவால்கள் உள்ளன. தமிழைப்பொறுத்தவரை பின்வரும் பிரச்சினைகளை எதிர்கொண்டாக வேண்டும்.

1. தமிழுக்கான மொழி அட்டவணைகள்
2. தரநிலைகள்
3. எண்கள் மற்றும் பிற குறியீடுகள்
4. பல்வகைமைகள்

2000 இலிருந்தே அரசு ஆதரவுள் gTLDகள் வர்த்தக ரீதியில் தமிழகத்தில் நடைமுறையில் இருந்து வருகின்றன. இணையத்தைப் பயன்படுத்தும் தமிழர்களின் எண்ணிக்கை மில்லியன் கணக்கில் இருந்தாலும், இதற்கு ஆதரவு அதிகமில்லை. இந்தியர்கள் ஆங்கிலத்தில் நன்கு பரிச்சயப்பட்டிருப்பதால், தமிழில் திணைப்பெயர்கள் அவர்களுக்கு அதிகம் தேவையில்லை என்பதாகக் கருதப்படுகிறது. ஆனால் தமிழில் இணைய உள்ளடக்கம் அதிகரித்துவருகையில், செம்மொழி தமிழ் மீதான கல்வி பெருகிவரும் வேளையில் இதன் தேவை நிச்சயம் உணரப்படும்.

2002 இல், மல்ட்டிலிங்குவல் இன்டர்நெட் நேம்ஸ் கன்சார்ட்டியத்துக்கு, குறியீட்டு நிலைகள் தொடர்பாக, உத்தமம் அமைப்பு ஒரு முன்வைப்பை அளித்துவைத்தது. யூனிக்கோடு 3.2 இல் தமிழ் குறியீட்டு நிலைகள் தொடர்பான திருத்தங்களோடு இது தொடர்புப்படுத்தப்பட்டிருக்கிறது.

ICANN உருவாக்கியுள்ள IDN சோதனைக்களம் ICANN, தமிழ் TLD ஒன்றுக்கான சோதனைக்களத்தை அறிமுகப்படுத்தியிருக்கிறது.

- உத்தமம். அமை infitt.org
- தமிழ்நாடு. அரசு. பாரதம் tn.gov.in
- அரசு.தநா gov.tn

ICANN மற்றும் IDN

முதல் 16 IDN ccTLD கோரிக்கைகள் இப்போது ஏற்கப்பட்டு அறிவிக்கப்பட்டுள்ளன. இவற்றுக்கான எழுத்துச்சர மதிப்பீட்டு பணி விரைவுச்செயல்பாடு மூலம் முடிந்திருக்கிறது.

இவை பின்வரும் நாட்டுக்கானவை: எகிப்து, ரஷ்யக் கூட்டரசு, ஐக்கிய அரபு அமீரகம், சவுதி அரேபியா. இதன் பொருள் என்ன?

எழுத்துக்கள் DNS வேர்நிலை அளவில் ஏற்கப்படுவதற்கான கடைசி படிநிலையே எழுத்து ஏற்பு நிகழ்முறையாகும். இது இப்போது பயன்பாட்டுக்கு வந்துவிட்டது.

மேலும் 5 கோரிக்கைகள் ICANN இல் பரிசீலிக்கப்பட்டுவருகின்றன. விரைவில் கோரிக்கைகளின் எண்ணிக்கை அதிகரிக்கும்.

1. தயாரிப்பு (கோரிக்கை அனுப்பும் நாடு அல்லது பிராந்தியம்). எந்த IDN ccTLD வேண்டும் என்பதை முதலில் ஒரு சமூகம் தீர்மானித்தாகவேண்டும். அதை அதிகாரபூர்வமாக வழிநடத்துவது யார், எப்படி என்பது குறித்தும் உரிய துணை ஆவணங்களை உருவாக்கி, தயார்நிலையில் வைப்பது குறித்தும் முடிவு செய்யவேண்டும்.
2. எழுத்துச்சர மதிப்பீடு: மேலே கூறிய நிபந்தனைகளுடன் ICANN க்கான கோரிக்கைகள் தீர்மானிக்கப்படவேண்டும். IDN ccTLD சரங்களுக்கான தொழில்நுட்ப மற்றும் மொழியியற் தேவைப்பாடுகள் தீர்மானிக்கப்படவேண்டும். உரிய விண்ணப்பங்களை ஆன்லைன் அமைப்பு ஒன்றில், தேவையான ஆதரங்களோடு சமர்ப்பிக்கவேண்டும். அதற்கான முகவரி - <http://www.icann.org/en/topics/idn/fast-track/>
3. எழுத்து ஒதுக்கீடு: எழுத்து மதிப்பீடு வெற்றிகரமாக முடிந்தால், ICANN எழுத்துக்களை ஒதுக்கீடு செய்யும். அதற்காக ASCII யில் உள்ள ccTLDகளுக்குப் பயன்படுத்தப்படும் அதே ICANN IANA நிகழ் முறைப்பயன்படுத்தப்படும். IANA வேர்நிலை அமைப்பு நிர்வாகத்துக்கு எழுத்து ஒதுக்கீடு விவரங்கள் அளிக்கப்படும்படி கோரப்படும்

மூன்று தமிழ் இணைய முகவரிகள் முன்வைக்கப்படும்.

1. தமிழ் நாட்டில் இருக்கும் நிறுவனங்கள் ,
அல்லது தனி நபர்களுக்கு - **தமிழ்நாடு**
எ.கா: அரசு.தமிழ்நாடு, நலகம்.தமிழ்நாடு, முதல்வர்.தமிழ்நாடு,
2. உலக தமிழ் மக்களுக்கு என ஒரு அடையாளம் - **தமிழ்**
எ.கா பாரதி.தமிழ், தென்றல்.தமிழ், குறுஞ்சி.தமிழ்
3. வர்த்தக நிறுவனம் - **வணி**
எகா. விகடன்.வணி , இந்து.வணி, நக்கீரன்.வணி

முடிவாக

இணைய பெயர்களை பன்னாட்டுமயமாக்கும் முயற்சி வெகுவாக நடந்துகொண்டிருக்கிறது. பழைய முறையிலிருந்து புதிய முறைக்கு மாறுவதற்கு முறைமை தரப்படுத்தம் உதவும். இனி பல பன்மொழி பதிவாளர்கள் நிறுவனங்களுக்கும் அமைப்புகளுக்கும் தனிநபர்களுக்கும் பன்மொழி திணைப்பெயர்களை பதிந்துதருவார்கள். ஆங்கிலம் பொதுவாக புரிந்துகொள்ளப்படாத நாடுகளில் இது இணையத்தை அனை வருக்குமானதாக ஆக்கும். இந்த மாநாட்டின் போது தமிழ் இணைய முகவரிகள் பதிவு நடைபெரு கின்றன. மேல் விவரங்களிக்கு <http://www.universal-names.com>, மின் அஞ்சல் contact@kanithamil.com

A Tamil Web Portal Development with online Dictionary

A.Parimaladevi, III MCA,
Kumaraguru college of Technology, Coimbatore,
priyr.th.cu@gmail.com

A.Muthukumar, Professor,
Kumaraguru college of Technology, Coimbatore
amuthukumar@computer.org

1.0 Introduction

1.1 Purpose

The purpose of this document is to present a detailed description of the mytamilthai.com which is a Tamil web portal development that includes all the details in Tamil. It will explain the purpose and features of the website. This document is intended for both developers and end-users of the system.

1.2 Scope of the project

This site is going to develop mainly for the people who feel uncomfortable while browsing in English. This site includes the information about youth, jokes, sports, medicines and also information for women in Tamil. Instead of searching different sites user can search a single site for his/her entire requirement.

This site also includes the English to Tamil dictionary. User who needs to search the Tamil meaning for a particular English word can effectively use this site. Also if the particular word is not found and if the user knows the meaning, then he/she can add the new word. Administrator who has the rights to see the words in temporary database and add to the existing database of the dictionary.

Some pages of the site can be viewed by all people but certain pages are viewed by the user only after they register their name in this site. Also authorized user can advertise for the jobs they provide with admin assist. Also user can post their comments.

2.0 About index page::

An index page is the first page of web portal system. The site mainly consists of four main divisions. The division at the top is Header with the title and division at the bottom with privacy policy and conditions are displayed in all pages of Tamil web portal system. The next division is for menus which will be displayed only in certain pages.

The menus includes the following

- இளமை - youth
- பெண்கள் - women
- நகைச்சுவை - jokes
- விளையாட்டு - sports
- மருந்தகம் - medicines

The last division will be content division. In this, there will be some sub-divisions. Left side which includes tabs like

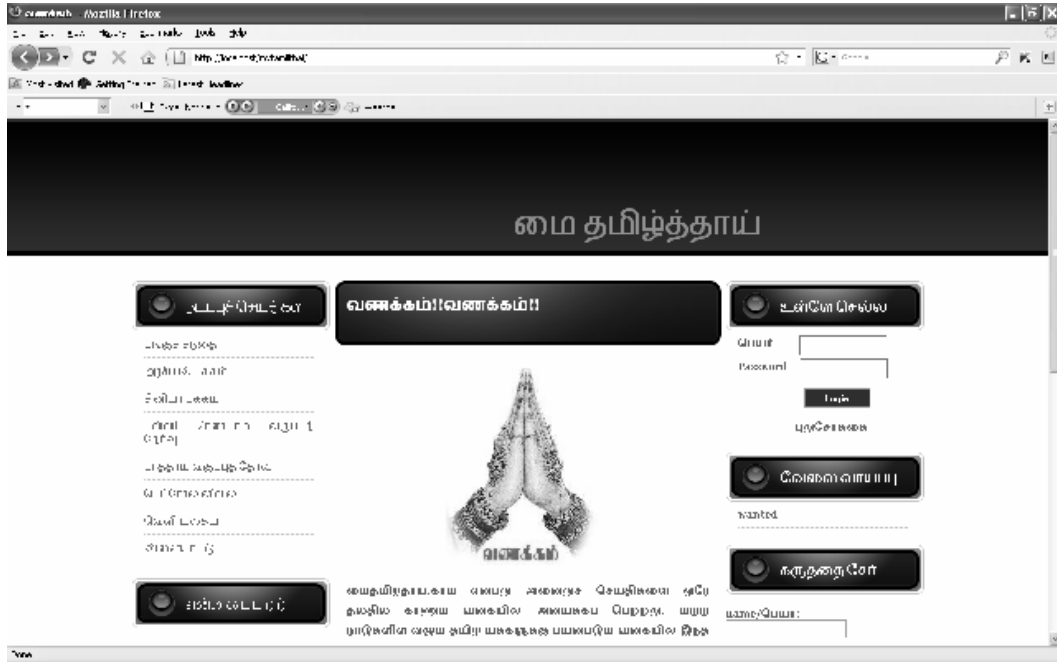
- நடப்புச் செய்திகள் - recent news
- என்னைப் பற்றி - about author
- மக்கள்கருத்து - user comments

In recent news, daily news will be updated for eg. About stock market, gold rate, politics, sports, etc. All the details about author will be displayed in next tab and at last if any user posts comments about any particular topic then with admin assist, the comment will be displayed.

The middle sub-division will be the exact content of particular topic. And the last sub-division includes the tabs like

- உள்ளே செல்ல - login
- வேலை வாய்ப்பு - jobs available
- கருத்தைச் சேர்க்க - post comments
- விசைப்பலகை - keyboard

To login, user needs to register in the site. The information about the registration will be stored in the database. All the jobs advertised will be displayed here. Every one can see all the information about that particular job by just clicking it. User who needs to post the comments, can post here but it will be stored in the database for admin to approve it to display in the site.



3.0 Admin login

Administrator who has all the rights on the system. As admin login, all the tabs displayed in index will be displayed with some extra tabs which includes

- அகராதி-dictionary
- புதிய சொல்சேர்க்க-add new word in dictionary
- புதிய வேலை-new job advertising

- வேலை வாய்ப்பு-advertised jobs
- விசைபலகை-keyboard

The menu will be enabled once admin login. As the menu contains the topics, the information regarding the particular will be displayed in the content page once we select the particular topic in that menu.

Online dictionary is available here. Tamil to English and English to Tamil dictionary is available with the description in both the languages. Admin add words to dictionary and it will be stored in database in Unicode format.

Admin who enters all the information about the job by the information provided by the job provider. Also already advertised jobs will be displayed.

4.0 User login

User after registering, user enters the site and views all the tabs displayed in index page with some extra tabs. User also can add words to dictionary but it will be stored in database. If admin find that word is irrelevant then he has the rights to remove the word from database so that, that word won't be displayed further unless again adding it properly.

Users who wish to post comments can post here. This is also stored in database for admin use.

5.0 Online dictionary

As described earlier, online dictionary will be displayed both in admin and users login. By typing English word we will get Tamil meaning as well as description in Tamil and the same way if we type in Tamil we will get English meaning with description in English.



6.0 Conclusion

MyTamilThai.com is developed to help all the people who need to know all the information at a single site. Here there is an availability of dictionary with English to Tamil as well as Tamil to English dictionary with glossary. So it is very useful for all people at different countries to the know the meaning for any word they wish. Recent updates are displayed and it will be updated daily.

Future enhancement: This project has been developed as a Master's project and is constrained by time. There is scope for extending the system as per the need.

Job providers advertise for the job availability now. Job seekers at present can view just the job advertisement and need to contact separately by them selves with the information provided by the job provider. Further it is going to extend by job seekers to post their resume for the advertisement displayed in the site through this site directly.

Also now there is an availability of English to Tamil and Tamil to English dictionary only. Further it is going to enhance in different language according to users need.

Also according to users authorized comments the site will be enhanced.

7.0 References

1. PHP 6/MySQL® Programming for the Absolute Beginner - Andy Harris -
 - a. Course Technology, 2009.
2. <http://www.w3schools.com/PHP>
3. <http://www.php.net/manual>
4. <http://www.actionscript.org>
5. <http://senthilvayal.wordpress.com>
6. <http://classroom2007.blogspot.com>
7. <http://www.webulagam.com>
8. <http://tamil.webdunia.com>
9. <http://www.thamilworld.com>
10. <http://www.adhikaalai.com>

Methods and Options for Videoconferencing in Relation to the Tamil Language in 2010

Eric Miller

On 21 February 2009, Chief Minister Kalaignar Karunanidhi in Chennai made a video call to launch the 3G (Third Generation) network services of BSNL, the Government-run telecommunications company. Today (May 2010), video calls on mobile telephones using BSNL's 3G network are being routinely made by members of the public in Tamil Nadu. Airtel's 3G network is also expected to be operational soon.

At this moment in time, as we are on the verge of the *Video Call Revolution*, it may be useful to look both backward and forward. This might help us to decide what to do with our new videoconference capabilities, as we enter the *Age of Videoconferencing* in earnest. Previously, I have written three articles about videoconferencing for INFITT -- in 2002, 2003, and 2004.¹ This article summarizes and adds to those.

Videoconferencing, video calls, video chat, or simply, *being able to see people as we speak with them through electronic devices*, has been on the horizon for many years. Two major ways of videoconferencing are: through one's computer, and through one's mobile telephone. In both cases, the hardware is increasingly coming with a video camera built-in. The videoconference camera is generally above the screen. In the case of mobile telephones, there are usually two cameras, one facing the user (for videoconferencing), and one facing away from the user (for optional use for recording still-images and video). Actually, the computer and the mobile telephone are converging, to produce *smart phones* -- and these are the mobile telephone models that tend to be videoconference-capable.

Skype is most commonly used on personal computers, but it can also be used through smart phones. Skype has brought videoconferencing to wider-than-ever-before public use. Other free video chat programs include those in Gmail and iGoogle, Microsoft's Windows Live Messenger, Yahoo! Messenger, and Apple's iChat.

Among online social networks (also known as, social media), Orkut is one of the most advanced in offering the video chat option to users. How will people with similar interests find each other to videoconference with? They could join *communities*, or become *friends*, *fans*, or *followers* of others, as many people already do on social media.

There are also programs such as Webex, which, for a cost, enable videoconferencing over the Internet, with sharing of files -- for text, electronic drawing, video, etc. -- in various windows.² In the 2004 and

1) "Videoconferencing and the Teaching of Tamil Language and Verbal Arts", <http://www.infitt.org/ti2002/papers/41EMILLE.PDF>, presented at TI2002, San Francisco Area, California, September 2002.

2) "Chennai and Videoconferencing: Videoconferencing for Performing, Teaching, and Discussing Tamil Language and Performing Arts", http://infitt.org/ti2003/papers/50_emiller.pdf, presented at TI2003, Chennai, August 2003.

3) "The 16 Oct. 2004, and 15 Oct. 2005, Webcasted-videoconferences for the Demonstration and Discussion of Children's Tamil (and Other) Songs/Chants/Dances/Games, and Methods of Teaching and Learning Spoken Tamil Language", <http://www.storytellingandvideoconferencing.com/27.html>, in *Min Manjari*, INFITT's e-Journal, December 2004.

² <http://www.webex.co.in>

2005 Chennai-Philadelphia videoconferences I facilitated, we used a simpler (and lower-quality) method for showing Tamil text: we projected it onto the back wall of our room in Chennai (Figure 2). In other videoconferences, I have placed the Tamil text next to the image of the speaker, like in a comic book (Figure 1). The development of ways of producing *instantaneous visual translation sub-titles*, or *captions*, is going to be an important part of videoconferencing's future. This will require *voice recognition technology* (from spoken to printed words), and *automatic translation technology* (from one written language to another).

Microsoft's *Natal* system will feature motion-sensing technology -- not just for game-playing, but for operating the computer in general. Thus, the *Gesture Revolution* is upon us, in which a crucial input device is a video camera.³ Based on what I have seen in the college students I teach in Chennai, many young people today will not rest until they can *play games via videoconference in social networks on mobile telephones*.

Regarding high-quality types of videoconferencing in Tamil Nadu, India, and beyond:

Early videoconferencing in India (in the 1990s) tended to occur via dedicated non-Internet ISDN lines (three lines together yield 384kbps). However, ISDN lines cost by the minute, and are quite expensive to use -- and are even more so when a bridge is needed, for connecting more than two partners in a videoconference.

The global development is now toward videoconferencing via very high-speed Internet. For example, Reliance -- whose videoconference rooms in their over 200 Reliance World stores across India have led the way in making high-quality videoconferencing visible and available -- previously offered only dedicated ISDN-line videoconferencing; now they also offer Internet videoconferencing.

Access Grid is an "ensemble of resources" that enables many sites -- using interactive multimedia and appearing on multiple screens -- to participate in a videoconference⁴ (Figure 4). Polycom, Cisco, LifeSize, and Tandberg are among the videoconference companies that offer *telepresence* (presence, from a distance), which involves life-size high-definition images of people, simulated eye-contact, and minimum delay-time -- replicating the experience of physically-present meetings as much as possible.

In India, three Governmental entities that are involved with developing videoconferencing -- or the connectivity systems that enable videoconferencing -- are 1) ERNET (Education and Research Network)⁵; 2) NIC (National Informatics Centre)⁶; and 3) CDAC (Centre for Development of Advanced Computing).⁷

NRENs (National Research and Education Networks) are playing increasingly important roles in Internet development in many developing countries. An NREN is a specialised internet service provider dedicated to supporting the needs of the research and education communities within a country. ERNET (Education and Research Network) is India's NREN.⁸

³ "Now, Electronics That Obey Hand Gestures", <http://www.nytimes.com/2010/01/12/technology/personaltech/12gesture.html>, New York Times, 11 January 2010

⁴ <http://www.accessgrid.org>

⁵ <http://www.eis.ernet.in>

⁶ <http://home.nic.in>

⁷ <http://www.cdac.in>

⁸ <http://www.eis.ernet.in>

Internet2 is a USA-based networking consortium.⁹ The technical standards and connectivity that Internet2 involves are very important factors in the global development of the Internet. Founded in 1996 by members of the education and research community, Internet2 provides

both leading-edge network capabilities and unique partnership opportunities that together facilitate the development, deployment and use of revolutionary Internet technologies... Internet2 brings academia together with technology leaders from industry, government and the international community...and promotes collaboration and innovation...¹⁰

Internet2 has a section relating to Emerging NRENs.¹¹ One of the Emerging NREN groups is the South Asia Special-Interest-Group (SA-SIG).¹² SA-SIG's mission is to help to facilitate high performance networking in South Asia. SA-SIG's e-mail list -- <southasia-interest@internet2.edu> -- is very much worth joining.

Two annual videoconference showcase events based in the Internet2 community are the Megaconference,¹³ which began in 1999; and the Megaconference Jr,¹⁴ which is especially for/by/with schoolchildren, and which began in 2004. Megaconferences are marathon events (up to 12 hours) that are webcast live to a global audience. They are composed of numerous brief videoconferences, each between up-to-six parties. I participated in the 2005 Megaconference, and in the 2006 Megaconference Jr (in both instances along with Tamil children, and exploring children's songs and language learning; from TENET's IIT-Madras facility). Reliance personnel have participated in a number of Megaconferences.

TEIN3 is another network that enables high-speed videoconferencing and assists in facilitating interesting collaborations.¹⁵ TEIN3 is the third generation of the Trans-Eurasia Information Network. With direct connectivity to Europe's GÉANT network, TEIN3 offers researchers and educators in Asia-Pacific a gateway for global collaboration with their peers in Europe and other parts of the world. India connected to TEIN3 in March 2010 (through ERNET). Sri Lanka connected to TEIN3 in April 2010 (through LEARN).¹⁶

An entity to keep in mind in relation to high-speed international videoconferencing for education and *development* is the Global Development Learning Network (GDLN).¹⁷ The GDLN, which is coordinated by the World Bank, is "a partnership of over 120 institutions in over 80 countries that collaborate in the design of customized learning solutions for people working in development".¹⁸ A regional association of GDLN, is GDLN Asia Pacific (GDLNAP).¹⁹ Presently there is one GDLN site in India, in New Delhi.²⁰ It might be good to seek to have a GDLN site in Tamil Nadu also. A GDLN site typically features a large room equipped with top-notch videoconferencing facilities.

⁹ <http://www.internet2.edu> .

¹⁰ <http://www.internet2.edu/about> .

¹¹ <http://www.internet2.edu/international/index.cfm> .

¹² <http://southasia.indiana.edu> .

¹³ <http://www.megaconference.org> .

¹⁴ <http://www.megaconferencejr.org> .

¹⁵ <http://www.tein3.net> .

¹⁶ Sri Lanka's NREN is LEARN (Lanka Education and Research Network), <http://www.ac.lk> .

¹⁷ <http://www.gdln.org> .

¹⁸ <http://www.gdln.org/about> .

¹⁹ <http://www.gdlnap.org> .

²⁰ TERI Distance Learning Center, New Delhi. <http://www.gdln.org/about/locations>.

Teaching Tamil language via videoconference -- on computers and mobile telephones -- could be a very important field. Tamil Nadu could be a world leader in developing language teaching in general by videoconference. These services should be available via Skype or similar programs, 24 hours a day.

At present, Tamil language-learning materials and instruction are available on webpages such as *Web Assisted Learning and Teaching of Tamil*.²¹ Such asynchronous learning processes could also have an optional synchronous videoconference component. The videoconference language-practice lesson-plans could be coordinated with the lessons on the webpages. The on-line tutors -- or language-practice partners -- would need to be recruited, trained, and put in contact with clients. This would involve a lot of work. It could be done as a business, an NGO, and/or an educational project, possibly subsidized by a government. In any case, it should be done, as a way of preserving, developing, and globalising the Tamil language. The early pervasiveness of the English language on the Internet is fading. Other languages are now also entering cyberspace, especially as the audio and video options are becoming more available and convenient.

My dissertation²² recommends three techniques for teaching language via videoconference: *Question-and-Answer Routines*, *Repetition with Variation*, and the *Simultaneous Saying and Physical Enacting of Words*. My research shows that these are prominent elements of Tamil children's songs/chants/dances/games - activities that very likely facilitate the acquisition of spoken language.

Question-and-Answer Routines place language-practice in the interactive context of human relationships, whether one participates in a routine as oneself or as role-playing other characters.

Repetition with Variation gives the learner a sense of control and competence. If only one aspect of a sentence is modified, the learner can still hold onto the grammatical structure of a sentence. Variations can include changes of tense, and substitutions of words (substitution drills); and going from a positive to a negative statement, or going from a command to a question (transformation drills). Repetition with variation is a key aspect of the modern language-teaching approach, the Audio-lingual Method. Systematic and methodical learners tend to especially enjoy this approach.

The Simultaneous Saying and Physical Enacting of Words utilises the entire body -- not just the brain and mouth -- in the language-learning process. The modern language-teaching method, *Total Physical Response*, is based on this idea.

These three practices are especially good for videoconference language teaching-and-learning, because in the course of videoconference communication it may at times be difficult to make out what a distant person is saying, and these practices makes interpersonal verbal comprehension more likely.

We also found that playing with puppets can add fun to videoconference language-practice. This additional level of mediation and role-play, which enables people to interact with each other indirectly, seems to relax people and take some pressure off them (Figure 3).

Tamil Nadu should take a leadership role in developing ways for many aspects of its culture to be shared via videoconferencing. Training, interactive performance, and discussion about the various

²¹ <http://ccat.sas.upenn.edu/plc/tamilweb> , by Dr. Vasu Renganathan.

²² "Ethnographic Videoconferencing, as Applied to Songs/Chants/Dances/Games of South Indian Children, and Language Learning", <http://www.storytellingandvideoconferencing.com/280.html> , PhD dissertation, Folklore Program, University of Pennsylvania, 2010.

aspects of culture could all occur via videoconference. This could be done in part in coordination with the State crafts organisation, *Poompuhar*; and the annual folk performing arts festival, *Chennai Sangamam*.

Videoconferencing in the classroom is a huge field in the USA and Europe, and students and teachers there are very eager to videoconference with their counterparts in exotic places like India.²³ Videoconference interviews with tradition-bearers are increasingly being held in Folklore and Social Studies classrooms.²⁴

Videoconference interviews for employment, and for admission to academic programs, are also becoming commonplace.

Chennai and other cities in Tamil Nadu need *teletoriums*: halls equipped with large screens and videoconferencing facilities. Setting up videoconference systems in halls for single events is too time-consuming and risky. A single teletorium could be used by numerous academic institutions and other groups.

A word should be said about a world-famous experiment in bringing videoconferencing (and general computer and Internet use) to the Tamil countryside, and to other rural areas in India. In the early 2000s, Dr. Ashok Jhunjunwala -- leader of TENET (Telecommunication and Computer Networking Group, Depts of Electrical Engineering, and Computer Science and Engineering, IIT Madras) -- helped to develop SARI (Sustainable Access through Rural India),²⁵ which was serviced by n-Logue Communications Private Ltd and other companies. While it seems that this project has for the most part proved non-sustainable,²⁶ Dr. Ashok Jhunjunwala continues to champion the idea of videoconferencing for people in the Indian countryside. It may well be that such people may achieve the ability to videoconference through mobile telephones before they achieve it through desktop computers. In any case, there remains a place and need for NGO and Government support for rural and economically-disadvantaged people to utilize videoconferencing for educational, employment, cultural, and other applications.

Apollo Hospitals is one among a number of hospitals in Tamil Nadu -- private and public -- that have strong tele-medicine components, using videoconferencing in the diagnosis process, for staff training, and for other applications. Sad to say, but difficulties for the environment (such as the ash cloud that recently floated above Europe), and inconveniences for travel, tend to cause booms for videoconferencing. Videoconferencing can be thought of as a *green activity*, in that it can reduce the amount of petrol needed for travel, and it can make it unnecessary for people to travel through delicate natural environments. However, videoconferencing should not be seen as a substitute for physically-present communication -- it is simply a different type of communication, with its own strong and weak points.

²³ *E-School News: Technology News for Today's K-20 Educator*, <http://www.eschoolnews.com/2008/05/27/internet2-expands-schools-possibilities> . The *Education Initiative of Internet2*, <http://www.internet2.edu/k20> , <http://www.internet2.edu> . The *Consortium for School Networking*, <http://www.cosn.org> .

²⁴ "Conducting Interviews via Videoconference", <http://www.afsnet.org/sections/education/Spring2008/Feature.html> , in the *Folklore and Education e-Newsletter* of the American Folklore Society, Spring 2008.

²⁵ <http://edev.media.mit.edu/SARI/sari-pilot-new-2.htm> .

²⁶ "Sustainability Failures of Rural Telecenters: Challenges from the Sustainable Access in Rural India (SARI) Project", <http://itidjournal.org/itid/article/viewFile/309/141> .

Years ago, people could often be heard in Tamil Nadu's browsing centres, singing Tamil cinema songs to distant others. Browsing centres have decreased, as Internet connectivity in the home, school, and office -- and on portable communication devices -- has increased. However, Tamil Nadu has been a world leader in bringing tribal, folk, and classical performing arts into the realm of cinema. Now these arts -- along with the teaching-and-learning of the Tamil language itself -- should be brought into the realm of videoconferencing, to further enable the sweet sound of Tamil to be heard and spoken around the globe.



A videoconference featuring electronic drawing and Tamil words (with transliteration and translation). (Figure 1.)



The words of a children's song are shown on the Chennai side of the Oct. 2004 Chennai-Philadelphia videoconference. (Figure 2.)



A conversation through puppets. From the Oct. 2004 Chennai-Philadelphia videoconference. (Figure 3.)



An Access Grid videoconference. (Figure 4.)

Eric Miller <eric@storytellinginstitute.org> is Director of the World Storytelling Institute (based in Chennai); and is Assistant Professor of Story and Storytelling at ICAT, the Image College of Animation, Arts, and Technology (also based in Chennai).

Thin Client and Server Based Computing to Provide Integrated School and Class Room Management System in Malaysian Tamil Schools

Saminatha Kumaran Veloo, Saravanan Mariappan,

Nexus IT Solutions, Malaysia,
e-mail: infonexusonline@gmail.com

Abstract

Most of the educational institutions now have extensive information and communications technology (ICT) in place. The cost of supporting, upgrading and replacing this equipment to provide a robust infrastructure for teaching and learning is increasingly onerous. This brings into question whether alternative network architectures, such as Thin Client computing could provide the required level of functionality with lower long- term costs and/or other benefit.

This paper addresses the use of thin client technology to provide optimum and cost effective solution for IT infrastructures which have the ability to integrate class room management application and school network management application with reliable and stable solution at minimum maintenance cost. The proposed thin client technology will be able to provide effective and secured centralized server based solution of schools with class room management systems integrated.

KEYWORD: *Thin Client, school and class room management, server based computing*

1 Introduction

This paper addresses the key area of institutional concern for the education sector, that of delivering effective and efficient school and class room management system in a flexible, secure, and accessible way to students in Malaysian Tamil schools. The system will adopt the thin client technology linked with centralized server to implement school and classroom management.

The proposed system will have secure integration with other key educational systems (e.g student records, module registration, and examination scheduling conducting trial exams and distribution of teaching materials), which will be delivered via network services and centralized server technology.

Thin client technology offers major advantages over conventional PC-based class room systems in terms of scalability, economy and sustainability. It can also offer additional flexibility in the range of school-teaching material and multimedia content that can be delivered without needing elaborate installation procedures or additional software to control the security, access and database.

2 Integration of the System

The system integration of Thin client based School and Class room management system with open source technology consists of:

- i) **School management System** software designed to automate a school's diverse operations from classes, exam to school events, calender and to create powerful online community, by bringing parents, teachers, and students on the common interactive platform.
- ii) **Local Centralized Server Based Technology** that consists of storage Server well equipped with database using Linux Edubuntu server platform.
- iii) **LAN/WAN network system** which allows thin client terminals to access the centralized server remotely via PXE LAN booting technology. This allows freedom for user to access their desktop from any thin client terminals within the network coverage.
- iv) **Class room management System** provides teachers with ability to instruct, monitor and interact with the students either individually, as predefined group or to the overall class.
- v) **Thin Client Server** provides solutions for central deployment, configuration and management of thin clients and users connections.
- vi) **Thin Client Terminals** that consists of motherboard, display devices, keyboard and mouse, without any preinstalled operating system and hard disk.

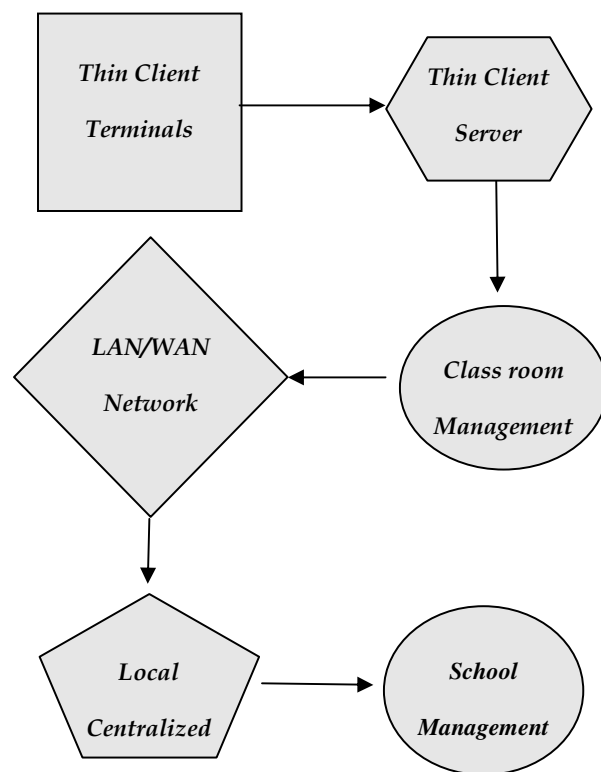


Figure 1 show the process flow of the system

Figure 1: System Integration of class room and school management system

2.1 School Management System

The school Management System has something for everyone related directly or indirectly with the school and teaching environment. Some of the key advantages to schools and educational institutions are:

- Easy performance monitoring of individual teaching modules.
- Automated and quick report generation along with process turn around time.
- Centralized data repository for trouble-free data access.
- Authenticated profile dependent access to data.
- User friendly interface requiring minimal learning and IT skills.
- Design for simplified scalability.
- Elimination of people dependent processes.
- Minimal data redundancy.

Some of the advantages to parents are:

- Frequent interaction with teachers.
- Reliable update on child's attendance, progress report and fee payment.
- Tracking of homework assigned by teachers to their child.
- Prior information about school events and holidays.
- Regular and prompt availability of school updates such as articles, discussions forums, image gallery and messaging system.

2.2 Class Room Management System

Advantages to teaching mechanism:

- Automated student attendance.
- Computerized management of marks and grades.
- Timetable creation in advance.
- Homework assignment to students and approval.
- Efficient and effective interaction with parents.
- Access to forum common to students and parents.
- Access to own and students attendance.
- Power on, power off, Reboot and Login to class room computers remotely.
- Broadcast messages to groups or all network users in seconds.

Some key benefits are:

- Enhanced interaction with teachers, parents and peers.
- On line submission of homework.
- Access to their attendance, timetable, marks, grades and examination schedule.
- Liberty to publish articles and views, and participate in discussion forums.
- Freedom to browse through library books catalogue and identify the book(s) to be issued.
- Prior information about school events and holidays.

2.3 Thin Client System

- Thin client is a general term for a device that relies on a server to operate.
- Thin client has display device, keyboard with mouse and basic processing power in order to interact with the server.
- An ideal thin client device contains no hard drives and CD or DVD-ROM

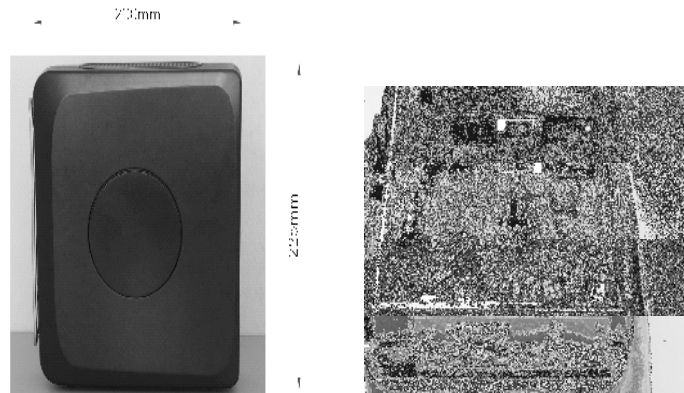


Plate 1: Ideal thin client architecture

3 Advantages of the Technology

By using thin client technology rather than standalone PCs, it is possible to deliver a wide range of computer based educational and examination materials while restricting other resources that are usually accessible to the students if conventional PC system is to be used. With conventional PC based technology, it is difficult to prevent access to the Internet, chat services, mobile devices such as USB drive, documents previously stored by other students etc., which could allow simple cutting and pasting of answers into the assessment or exam sheets by students with thin client technology in place. It is simple for an administrator to disable USB port on thin client terminals for the duration of the assessment or examination time, thus further limiting the ability for student's accessing disallowed information to assist them in the assessment or examination.

Another major attraction of the thin client technology for assessment purpose is that it is very resilient, given the fact that they have no software or moving parts. Therefore there are unlikely to be an issue when the assessment are not been delivered due to faulty desktop devices. This causes unnecessary pressure on the affected student and the additional works involved to the invigilator.

The issue of ensuring that PC's have the appropriate software available also affects PC's which are located in teaching spaces. Traditionally such PC's are left switched off when not in use which means that any automated software updates tend to fail or, worse, try to start when a teacher turns the PC on for a class. This can lead to anti-virus software not being updated, operating system vulnerability not being patched etc. the start up time of a PC system also causes difficulties, when a lecturer arrives in a class room, there will be about 8~10 minutes start up time for the PC and to get the necessary software up and running; if any updated needed to be done this could delay the start of the class. Using thin

client technology there is no need for the software updates and no need to worry about viruses. The user will always get the appropriate version of all the software via central server. The new upload of teaching material will be ready for teaching immediately as the student or teacher starts the class. Figure 2, shows the flow between teachers-students-parents-school in the school and class room management system.

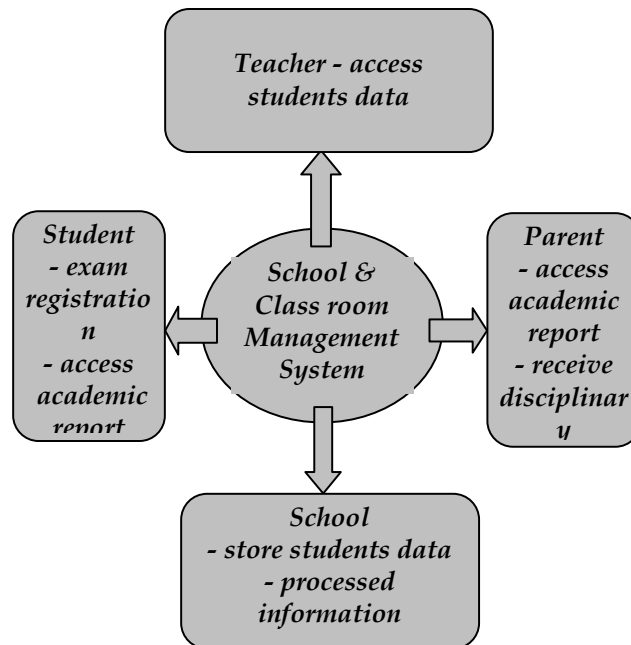


Figure 2 : Data Flow between teachers-students-parent-school in the school & class room management system

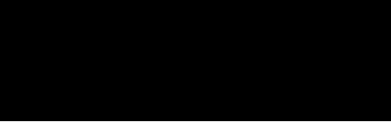
4 Conclusion

The school and class room management system with thin client technology have not been fully integrated in the learning process in Malaysia. In last decade, the complexity of existing desktop machines, the capital investment needed for wide area network (WAN) access and lack of educational resources and multimedia content have prevented the potential of thin client based solution become reality.

Recently the convergence of community, business and government organizations in favor of client technology, have started to produce changes in education system. With the use of thin client technology, the teaching system will now can look forward into a new age of centrally manageable teaching technology, with equal access to information will be given to all students regardless of their background and geographical location. Rural students will be given full access to information and knowledge with tremendous reduction in communication time and infrastructure cost, knowledge can be shared with anybody from any part of the world for free. Coupling the centralized server with thin clients and the power of cloud computing, Tamil schools in Malaysia, soon will become the community information hub. The students who benefited with this technology will one day become knowledge based skill workers which will directly uplift the living standard of Malaysian Indian.

8

**கணினி வழி தமிழ் எழுத்து
உணரி செயற்பாடுகள்**



Embedding Co-Features in 'Ocr-Friendly' Fonts will go a Longway in Machine Reading of Texts

N D Loga Sundaram

26/15 Kutchery Lane Mylapore, Chennai, India.

Cell : 091 044 9283244798, Alt. 9283772110

selvindls61@gmail.com,

Preamble

The Regal superiority of digital world, crowned by its hyper speed-error-free signal transmission i.e. its unparallel dynamics in information handling. Because IT being made as retrofits to faculties handled by humans it gained popularity very quickly. Today they are ubiquitous products in the hands of informed community.

Every one of us knows well that innovations in IT field are born worldwide, every minute, both in hardware and in software. By each of its million facets, they get deep-rooted mm by mm, in minute-wise human activity around the globe, recognizing that 'Continuous Improvement ' is one of the core theme for remaining in forefront in any business.

Playing field

The overwhelming, inundating expansion of human activity in the digital virtual universe, i.e. in the interface between the human (with their archives), to the current digital machines are limited to, Vision and Sonics. Applications connected with vision and Sonics are already in use, over decades, one after another - Vision to digits and back, Sonics to digits and back, cross platform between Sonics and Vision thro digits.

Archives- CORE instrument in human advancement

As pointed out elsewhere the whole spectrum of human advancement is sourced by the creation of huge voluminous Banks of Knowledge in visual/written format, which bypasses time and distance in human civilization. It is the vessel that accumulates and vends knowledge so that it can be continuously reviewed updated and suitably edited (evolution) to their customized needs on date. Archived written language (vision) is prone less to corruption and distortion than a spoken (Sonics) one and hence they survived perfectly such a long way since Stone Age until this digital era.

Primary in-put medium to human knowledge

Modern civilized humans get their knowledge mostly through sight and sound, both in real world as well from archives, thro 'language' that is called Education'. The other faculties as taste, smell and feeling play lesser part quantitatively which still remains primitive and mostly limited to tangibles in front as non-virtual.

The apparent necessity of quicker machine reading

From times immemorial humans devised machines for many of their day-to-day activities by inherent laziness and to thrive with comfort in competitive world out of their disturbing ambience. This is true in digital world also, which plays an arterial roll on date in 'human resource'. By nature I.T. is one of the platform erected for high altitude launch of mechanization and automation.

In the present world of communication explosion and consumerism we see people everywhere like to, rather forced to, have anything done quickly and easily, seldom even by fancy mania. The digital gadgets ranging from tiny hand held devices crowded with numerous features to mammoth net search engine servers, are the gadgets white collared people have interface these days, before middling with their bowl of soup. Next to TV and consumer durables, cunning high-tech assisted marketing techniques paved way for deep penetration of desktop PCs with all of its army of useful peripherals and the gadget of the day the cellular Mobile. Hence the universe of digital gadgets is in reign today and its influence extends well beyond informed community

Digital world and its Contents

Aptness and forerunner ship in handling of contents whether in a simple and humble home paperless-office or goliath popular commercial portal; it can be achieved by size of handled quantity as well as by timely presentation. Hence a self-employed free-lancer or a web master, he/she has to handle their contents, as ease and efficient as possible. In most cases, when the source is not from ready-in-the-shelf digital form, they get them digitized through their peripherals. In most of the cases, they are compiled from contents of several other authors, in print media, which warrants first degree of conversion.

Anyone who wares the shoe of a web master or a mere single document initiator they have initial digitization hardships and hiccups by virtue of time and strain it takes to key-in the contents thro' human eye-fingers route. Therefore they always opt for an easier machine input process if available replacing the laborious ubiquitous keyboard route which causes fatigue, back pain and eye irritation from not so ergonomically designed work place.

Contents in text format

The contents available in texts format instead of an image format is preferred for memory size and more particularly when it has to be subjected to an analytical process as search compare and manipulate with mathematical or logical operators. Compiling is one of the core processes in any creation and hence pick a particular content or theme from the group is the necessity. Therefore text formats are first and foremost choice.

'Always stands in top'

Hence the case of machine reading from print media stands on top of wish list of authors. All the attempts focused to this theme, will be appreciated and hence they command huge market. By virtue of this commercial proposition, several OCR products were brought out already and serve well for about a decade.

Again a review of the need and worth of machine digitization

Replacement of human activity in archiving information thro' digital machine paves way for

1. Unbiased,
2. Quicker,
3. faultless and
4. closely managed

handling of information which leads to omnifarious economics. It shall even be the necessity in future high speed lanes in E-management.

MICR is the forerunner for OCR

Machine reading of printed characters are not new while it is about 30 years old, which is well before the advent of popular use of digital data with computers. It was introduced as magnetic character recognition in bank cheque nos known as MICR cheques. It was devised to avoid human failures and helps in speedy unbiased disposal by using a special kind of ink and a process to suit the machine reading along with safety. Initially it was limited to serial number of cheques (numerals)

Now, design of same numerals of those safety cheques got modified, with simple distinct features in its normal printed glyphs. On date, one can see combination of thinner and thicker strokes or micro sized squares placed at a particular zone with disparity in each numeral's grapheme for providing easily traceable differentiating features between each of the characters in the set.

Bar code Reading

Then the digital era of barcode characters with 59 white and black bars was introduces and it is still a popular true workhorse at every marketing/vender outlet stores. Bar-coded price/product tags embed numerous particulars, which have bearings, both for seller as well as purchaser. As the bar code is devised purely for machine reading process, it is not suitable for normal/average human cognizance.

2D Checkerboard type tags

Now another innovation one can witness is a kind of tag (or nameplate) replacing bar-coded tag. It is a printed square box puzzle checkerboard matrix, which has rows of black and white optically, coded fields. By its 2D nature it contains more data than a barcode and evidently needs different reading writing devices to suit the system in design.

RFID

High volumes of digital activity also bring newer problems and we know about the necessity of safety of contents handled is an indispensable factor in any commercial/legal propositions.

Mechanization in identification or authorization using a radio frequency signals and embedded 3D micro devised chips on to products are in use today for remote sensing at a distance which already started gaining popularity.

Currencies

We know that the safety features are inlaid in Currencies of most of the nations of the world have features for non-human cognizance serves automated or machine recognition.

Embedded Features in Ocr Friendly Fonts

By applying the same basics in MICR, i.e. introducing an inlaid easily traceable secondary features to font characters/glyphs, we can bring about automation/machine reading of contents along with visibility to human eye in print/display media. They shall be embedded with distinct features within the fields of a font character, which can be both in positive or negative videos or combination of both.

Even any of the other durable and safe means, in print friendly high-tech means such as magnetic electrostatic chemical and radioactive are not ruled out. Of course in those case the term 'OCR' itself may have to be substituted.

Since OCR product with these new secondary features shall be designed with sensing mechanism focused to the particular zone of distinction in the field of a graphemes without wasting resources on confusing, inert and noise zones /feature, it is easier to design while the output shall be quicker and errorless.

Another advantage in introducing secondary features into graphemes in print media is that it will relieve many problems in design of OCR product for variations in colour, glyph size, style etc. while giving a font and print designer a free hand to their varieties as secondary features can even be made independent of normal optical visibility.

Forthcoming Contents

Being a newer introduction, it is for forthcoming contents in print media and not for existing contents in printed archives.

Early Deployment

To provide full fledged scope of freedom to font authors for their individuality and to cater their own specific vending zones no suggestions initiated at this stage other than a request to every one in digital planet to look into the scope for innovation and economics. This can cut short the take-off time.

Standardisation

For popularization of any product, standardization and universality is a natural prerequisite leading to general economics and convenience. This should be true for OCR products too. Therefore the developers of OCR-FRIENDLY fonts embedded with unique attributes can come to a common platform for standardizing their process and products so that plurality of vendor specific products has to be avoided since customer specific interests will always prevail upon in long term business.

I P Rights

This basic innovative theme of having a secondary feature in each of font grapheme cannot have intellectual property rights because it is merely a generalized idea re-presented. This lacks newness/state-of-art creative intelligence. But the design rights over a fonts' looks shapes and styles along with process/design of producing a secondary easily traceable visible or other feature remains with the author. Creation of distinct attributes between characters of a set/group when achieved through an

innovative **state-of-art** process in the components of printing and reading like font printing ink or else and in the nature of medium over the print is made, the IP rights can still be enforced.

Tamil and Universality

Though the embedding co-features for OCR-FRIENDLYNESS by virtue of universality, it shall be applicable to fonts/glyphs/graphemes of every language of the world, but when introduced in our Tamil as a pioneering work we can attract admiration from digital global villages. There will side effects and problems to be faced by Unicode Consortium in their monster Universal Character Set though their character-encoding scheme are not connected with shapes and looks and other features of fonts and graphemes.

Conclusion

I hope by virtue of its rationale printed graphemes embedded with secondary co-features will soon shine with its usefulness in the digital sky and contents with this innovation employed will be flooded with richness of every language, including Tamil, as an avalanche.

The apparent fallout in having maiden digital Tamil contents initially from OCR friendly prints they can easily be converted/used in any other output mode as Sonics or Vision. If we find in the roadmap of having speech to text and text to speech applications until the initial sonic input route is less efficient and more cumbersome, the OCR route can be useful.

Creation of annotated Tamil handwritten word corpus for OHR

Nethravathi B, Archana C P, Shashikiran K & A G Ramakrishnan

MILE Lab, Department of Electrical Engineering, IISc, Bangalore, India.

{nethra, archana, shashikiran, agr} @mile.ee.iisc.ernet.in

Abstract

Annotated datasets form a critical aspect in the development of robust technology for handwriting recognition and can be used for comparing results of different techniques used by various research groups. This paper describes the efforts at MILE lab, IISc, to create a database for the design and development of Tamil Online Handwritten Recognition. 100,000 words have been collected from 500 writers in Tamil, so that as much variations in writing style is captured. The data collected incorporated all the symbols (base characters, Indo-Arabic numerals, punctuations and other symbols). An annotation tool has been developed which helps the study of various styles of writing, stroke directions and presence of delayed strokes. Quality tags like class A, B, C etc has been assigned to the words accordingly. The annotated data is stored in a standard XML format defined by OHWR Consortium.

1. Introduction:

Databases are of great importance in any field of research, and handwriting recognition is no exception. A good database of handwritten data can be used to train and evaluate the performance of the recognition engine. Databases for scripts like Roman and Chinese already exist, whereas no such databases exist for Indic scripts. The database collected at MILE lab, IISc contains a comprehensive collection of words in Tamil, collected from many native Tamil people. Predefined word lists have been used to collect data, where the word list covers all the characters in the language. Here the focus is to develop a comprehensive database to support the development of a robust recognition engine. These databases facilitate comparison of different engines and also allow researchers to focus on recognition methodologies. A large database helps in removing bias of the engine towards particular styles of writing.

Tablet PC and G-Note have been used to collect data. The writer writes with an electronic pen on the electrostatic pressure sensitive writing surface of a Tablet PC or G-Note. The device captures the movement of pen tip on its screen in terms of x, y co-ordinates, sampled at equal intervals of time. It also captures the PEN_DOWN and PEN_UP information. The recognition is challenging because of varying styles of writing the same character. This paper describes how the database of 100,000 words has been collected from different schools and colleges, which involved major field work.

The collected data is annotated at the word, stroke group and akshara level using an annotation tool [2] developed by MILE lab. An akshara in Indian languages is a cluster of graphemes that need to be considered together to obtain the correct Unicode representation. Aksharas can be consonants (C), vowels (V) or a combination of them such as CV, CCV and so forth. The output of annotation is stored in the standard XML format [3] which was proposed by the OHWR consortium.

2. OHWR Consortium funded by TDIL:

A consortium made project was funded by Technology Development for Indian Languages (TDIL), Department of Information Technology, Government of India in January 2007 for research on online handwriting recognition. The project aims at developing Online Handwriting Recognition (OHWR) engines for Tamil, Kannada, Malyalam, Telugu, Bangla and Devanagari scripts. We at MILE Lab, IISc, are developing Tamil and Kannada engines. The academic partners of this project are IIT Madras, ISI Kolkata, and IIIT Hyderabad. The private and public industry partners are Learnfun Systems Chennai, CK Technologies Chennai and CDAC Pune.

3. Characteristics of Tamil handwriting:

Tamil compound characters (aksharas) are formed by graphically combining the symbols corresponding to consonants and vowel modifiers using well defined rules. Segmentation of words in these languages is more feasible than it is for English cursive writing as the characters are written separately without much overlap between them. In Tamil script, the majority of vowel modifiers are written as separate symbols and hence they are recognized separately.

4. Selection of complete constituent symbols:

Tamil: Tamil script comprises 313 characters. Of these, 12 are pure vowels and 23 pure consonants. Thus there are totally $12 \times 23 = 276$ consonant-vowel combinations. Apart from these, there are two additional symbols. The set of pure vowels in Tamil and its corresponding transliteration in English is depicted in Fig 1.

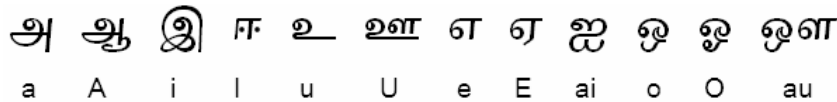


Figure 1. Set of Vowels in Tamil

We have established that only 155 symbols are required to represent all the 313 characters. The details are given below,

- 1) The vowel modifier for /A/ is depicted by a separate symbol and is written to the right of the consonant. Treating this vowel modifier as a separate class reduces the number of classes. A consonant ட /T/ combined with the vowel modifiers /a/ and /A/ are shown in two different rows of Fig 2.

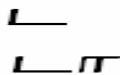


Figure 2. Consonants /Ta/ and /TA/

- 2) Vowel modifiers of /i/, /I/ and /u/, /U/ create new symbols when combined with the consonants. These new symbols are treated as different classes, thereby adding to the total number of classes. An example of this is shown in Fig 3.

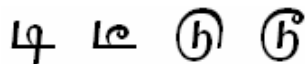


Figure 3. Consonants /Ti/, /TI/, /Tu/ and /TU/

3) The vowel modifiers of /e/, /E/, /ai/ are separate symbols written to the left of the consonant. These symbols are also treated as separate classes, further reducing the number of classes. Fig 4 shows an example of a consonant in combination with these vowel modifiers.

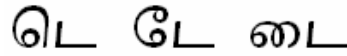


Figure 4. Consonants /Te/, /TE/ and /Tai/

4) The vowel modifiers of /o/, /O/ have two separate symbols which are written on either side of the consonant. The consonant combined with vowel /o/ will have the modifier of /e/ to its left and the modifier of /A/ to its right. Similarly a consonant combined with vowel /O/ will have the modifier of /E/ to its left and the modifier of /A/ to the right. Since these symbols are already handled separately, the number of classes reduces further. Example of a consonant combined with these vowel modifiers is shown in Fig 5

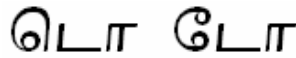


Figure 5. Consonants /To/ and /TO/

5) The vowel modifier /au/ also has two symbols with one written on either side of the consonant. The symbol to the left of the consonant is the same as the modifier of /e/ and the symbol to the right is the same as the consonant /La/. These two symbols are already handled separately, similar to case 4, which also causes a reduction in the number of classes. A consonant combined with vowel modifier of /au/ is shown in Fig 6.



Figure 6. Consonant /Tau/

Along with the characters, special symbols like full stop and question mark are also incorporated in the symbol list. It is to be noted that in modern Tamil script, Tamil numerals are rarely used. Hence these symbols are not included in our dataset. Hindu-Arabic numerals have been included, and treated as special symbols in our work. The words have been carefully chosen so as to represent all possible symbols used in modern Tamil script.

5. Data Collection for Tamil OHR:

5.1 Criteria for selection of acquisition devices:

The devices used for data collection are the Tablet PC and G-Note. G-Note is more suitable for field work as it is sturdy, affordable and easy to carry. It is also easy for the user to write on a G-Note as the feel is the same as writing on normal paper or pad. The data collected in G-Note is stored as .TOP files. Tablet PC is suitable for individual use. It is heavy and difficult to carry. Also since it is expensive, it cannot be used for field work. The TabletPC data is stored in .txt format.

These devices are shown in Fig 7. Predefined word lists have been used to collect data. A Tamil sample handwritten pages is shown in the Fig.8.



Figure 7. Tablet PC and G-Note 7000

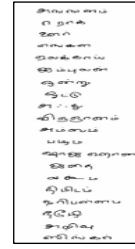


Figure 8. Sample handwritten Tamil Page

5.2 Selection of Writers:

The criteria for selecting writers for data collection was that the person should be a native writer of the language and one who is currently writing regularly. Students and teachers were primarily chosen for data collection as they write regularly.

6. XML Standard for Annotation:

The output of annotation is stored in a standard XML format which has been defined by the OHWR consortium [3]. This standard XML includes all the details about the data, such as the writer details, the device information, the number of pages and words. The words are truthed at the word level annotation. The aksharas and stroke groups are truthed at the character level annotation. All this information is stored in the XML. The XML also contains information about the quality assigned to each word, akshara, stroke group and stroke. This facilitates separation of Class A/good data from the Class R/reject data.

7. Annotation Details:

Once the data is collected, the first process is to do the word level annotation. The collected set has multiple words on each page; hence determined word boundaries are to be used to obtain the strokes of a word. In word level annotation, each word is labeled, using a tool developed by IIIT Hyderabad. The output is stored in a standard XML format defined by the OHWR consortium.

Next is the character level annotation, where the output of word level annotation is given as input. In character level annotation, words are separated into strokes, stroke groups and aksharas and they are labeled. Quality tags are also assigned to them based on the direction of writing, stroke order and validity of strokes. The output at this level is also stored in the standard XML format. This annotation at character level is performed using a tool developed at MILE Lab, IISc [2].

8. Quality labels for Strokes, Stroke groups and Aksharas:

The strokes, stroke groups and aksharas are assigned various quality labels based on the nature of writing. The labels are defined as follows:

Class A: Denotes words written correctly with the expected number of strokes and in the expected direction. They are automatically segmented correctly by the segmentation module. Based on the statistics of writings from a huge number of native writers, there are multiple sets of stroke sequences valid for many stroke groups.

Class B: Denotes words which require manual segmentation and stroke groups with 10% or less overlap.

Class C: Denotes words where two or more normally separate strokes are written as a single stroke or vice-versa. It also includes strokes with overlap greater than 10% and delayed strokes.

Class D: Denotes words with extraneous strokes or overwriting and strokes written in the opposite direction. However, the resulting stroke groups must have the potential to be properly recognized using offline features.

Class R: This is the reject class, containing wrong words and/or strokes for which the likelihood of recognition is very low.

9. Conclusion:

This paper describes how the database has been created for Tamil Online handwriting recognition. The process of creating a reduced symbol list, which includes all the basic symbols of the character set has been described. The focus is on the process of collecting data, the devices used the criteria for selection of writers and why the reduction in number of symbols is required. This paper also tells how the data can be annotated for further use by researchers.

10. Acknowledgment:

This entire data collection effort was funded by Technology Development for Indian Languages (TDIL), Department of Information Technology (DIT), Govt. of India. We thank Mr. Suresh, Mr. Rituraj, Ms. Chandrakala, Ms. Archana, Ms. Saranya and Ms. Sountheriya for their efforts which made this task successful. We thank Prof. Deivasundram (University of Madras), AVM Matriculation Higher Secondary School, Govt. Boys Higher Secondary School, Sulur, Presidency College Triplicane, Virugambakkam, Chennai and IIT Madras for contributing to the data set. We also thank Dr. Anoop Namboodiri for giving us the word level annotation tool.

References:

- [1] K. H. Aparna, V. Subramanian, M. Kasirajan, G. V. Prakash, and V. S. Chakravarthy. Online Handwriting Recognition for Tamil. Proc. 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR-9), 2004.
- [2] C. P. Archana, K. Shashikiran, and A. G. Ramakrishnan. A Stroke group to Word Annotation Tool for South Indian Languages. submitted to ICFHR 2010, Kolkata, Nov 2010.
- [3] S. Behle, S. Chakravarthy, and A. G. Ramakrishnan. XML standard for Indic online handwritten database. Proc. International Workshop on Multilingual OCR, Barcelona, Spain, 2009.
- [4] A. S. Bhaskarabhatla and S. Madhvanath. Experiences in Collection of Handwriting Data for Online Handwriting Recognition in Indic Scripts. 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, 26-28 May 2004.

Difficulties in developing OCR for Tamil documents and an Integrated OCR Solution

R. DhivyaBharathi

IT Department, Pre-final year student
Bannari Amman Institute of Technology
Erode - 638 401.
Email id: divyit@gmail.com

A. BalaMurugan

EEE Department, Pre-final year student
Anna University
Tiruchirappalli - 620 024.
Email id: rajiv.divy@gmail.com

Abstract:

In this paper, the major difficulties arising in the development of OCR for Tamil language is presented with solutions. We have proposed a new approach of using a Neural Network integrated with an Encoded Character String Dictionary. Multiclass Hierarchical Support Vector Machines (MHSVM), Hidden Markov Model (HMM) and Radial Bass Function Neural Network (RBFNN) are used for accurate character recognition.

As MHSVM has shown excellent character recognizing results due to a strong mathematical foundation, it is chosen for the recognition process. HMM is chosen for recognizing online written letters.

RBFNN is integrated with an Encoded Character String Dictionary and trained using the output of MHSVM and HMM for accurate character recognition. The three algorithms used here facilitate the recognition of Online, Offline and Printed Tamil characters which pave the way of developing an Integrated OCR. The main advantage of developing an Integrated OCR is to decrease the memory requirement and to increase the processing time. The proposed model is expected to be 90%-95% accurate.

Keywords: *Optical Character Recognition (OCR), Hidden Markov Models (HMM), Multi Hierarchical Support Vector Machines (MHSVM), Radial Basis Function Neural Network (RBF-NN), Encoded Character String Dictionary.*

1. Introduction

The common problem in all type of algorithms used in OCR is the set of letters in Tamil which look very much the same, except for some minor differences (Example: la and va). And we face more problems with the slanting letters. Straightening these slanting letters introduces more noise due to the digital nature of the pixels. If we keep them in their original shape, splitting them into

letters have to be handled carefully. These problems can be handled by looking at the previous letter or group of letters to decide the next letter using a dictionary of words. MHSVM provides excellent Offline and Printed character recognition. MHSVM requires little prior knowledge which saves a lot of time in training the neural network and minimizes risk. HMM is chosen for recognizing online written letters. Pen-up strokes are modeled using a left-to-right HMM and inter-symbol strokes are modeled explicitly in two states. This shows low performance in case of high lexicon size which can be eliminated using a

statistical method. The statistical tool can be provided by the MHSVM explained earlier. HMM also avoids the segmentation problem, which is extremely difficult and error prone.

2. Proposed OCR Algorithm

The accuracy or efficiency of OCR purely depends on the algorithm we deploy. The efficiency decreases when an algorithm fails to identify a character or if the algorithm detects an unrelated character. We have proposed a method where we can fuse two pattern recognition algorithms and evaluate the efficiency of OCR. Before fusing, the scanned document is preprocessed.

The steps in preprocessing involves

- 1) Histogram equalization and Gabor Filtering
- 2) Binarisation
- 3) ROI extraction
- 4) Region Probe Algorithm

2.1 Preprocessing Steps

This section describes the preprocessing steps of the scanned document in detail.

2.1.1 Histogram Equalization and Gabor Filtering

The scanned document is first applied to Histogram equalization and Gabor Filtering. Histogram equalization usually increases the local contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. Then the Gabor filter is applied to the scanned document obtained by the previous step by spatially convolving the image with the filter. A Gaussian function [24] multiplied by a harmonic function defines the impulse response of the linear filter, the Gabor filter. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function.

$$\text{Where, } x' = x \cos\theta + y \sin\theta \text{ and } y' = -x\sin\theta + y \cos\theta$$

In this equation, λ represents the wavelength of the cosine factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, Ψ is the phase offset, and γ is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function.

2.1.2 Binarisation

The binarisation process involves analyzing the grey-level value of each pixel in the enhanced image, and, if the value is greater than the global threshold, then the pixel value is set to a binary value one; otherwise, it is set to zero.

2.1.3 ROI Extraction

We perform morphological opening on the grayscale or binary image with the structuring element. We also performed morphological closing on the grayscale or binary image resulting in closed image. The structuring element is a single structuring element object, as opposed to an array of objects for both open

and close. Then as the result this approach throws away those leftmost, rightmost, uppermost and bottommost blocks out of the bound so as to get the tightly bounded region just containing the bound and inner area.

2.1.4 Region Probe Algorithm

After all the above process, the image is passed to the segmentation phase, where the image is decomposed into individual characters. For this we have used region probe algorithm.

2.2 Algorithm Fusion

Then we have fused two algorithms meaning that both the algorithms are taken into consideration. While fusing the two algorithms the following points are taken into consideration

- 1) If one algorithm fails to identify a character, another algorithm may support in identifying the character.
- 2) If one algorithm gives wrong character another may give a correct one.
- 3) The possibility for same wrong identification by both the algorithms is less.
- 4) If one algorithm gives wrong result the decision of choosing the correct result is done by neural network which is discussed later in the paper.

We have chosen SVM, HMM for the fusion and discussed in the rest of the section.

2.3 Hidden Markov Model (HMM)

Hidden Markov Models are suitable for handwriting recognition for a number of reasons [8]. The importance of HMMs in the area of speech recognition has been observed several ago [25]. In the meantime, HMMs have also been successfully applied to image pattern recognition problems such as shape classification [26] and face recognition [27]. HMMs qualify as suitable tool for cursive script recognition for a number a reasons. First, they are stochastic models that can cope with noise and pattern variations occurring in human handwriting. Next, the number of tokens representing an unknown input word may be of variable length.

Moreover, using an HMM-based approach, the segmentation problem, which is extremely difficult and error prone, can be avoided. This means that the features extracted from an input word need not necessarily correspond with letters. Instead, they may be derived from part of one letter, or from several letters. Thus the operations carried out by an HMM are in some sense holistic, combining local feature interpretation with contextual knowledge. Finally, there are standard algorithms known from the literature for both training and recognition using HMMs. These algorithms are fast and can be implemented with reasonable effort. Kundu and Bahl built an HMM for the English language [28].

However, they require the input word being perfectly segmented into single characters.

The Hidden Markov Model is a straightforward generalization of ordinary probability distributions to the case of randomly generated sequences of discrete or continuous-valued events. A discrete density HMM produces strings $O=O_1 \dots O_T$ of symbols from a finite alphabet $\{V_1, \dots, V_k\}$ while the continuous density version creates sequences of real-valued feature vectors $x \in IR^d$. The generation of the observable output $t O$ of the model is controlled by a doubly stochastic process. At each time instant $t = 1, \dots, T$ the model is in one out of N possible states $\{S_1, \dots, S_N\}$. The state q_t taken by the model at time t is a random

variable which depends only on the identity of its immediate predecessor state. According to this assumption the state distribution is completely determined by the parameters.

$$\pi = p(q_1 = s_i) \text{ and } a_{1,j} = p(q_t = s_j \mid q_{t-1} = s_i)$$

in other words, the vector $\pi = (\pi_1, \dots, \pi_N)^T$ of initial probabilities together with the $(N \times N)$ - matrix $A = [a_{i,j}]$ of transition probabilities form a first-order Markov chain.

The actual state sequence taken by the model serves as a probabilistic trigger for the production of the output sequence. The q_t themselves, however, remain hidden to an observer of the random process. According to a second model assumption, the probability distribution of an output symbol O_t (or an output vector, respectively) depends solely on the identity of the present state q_t ; thus, the distribution parameters.

$b_{j,k} = b_j(v_k) = p(O_t = V_k \mid q_t = s_j)$ of a discrete density HMM constitute a $(N \times K)$ probability matrix $B = [b_{j,k}]$. Consequently, the behavior of an HMM with discrete output is entirely specified by the cardinality N of the state space, the alphabet size, and a parameter array $\lambda = (\pi, A, B)$ of non-negative probabilities, obeying the $(1 + 2N)$ normalization condition

$$\sum_i \pi_i = \sum_j a_{ij} = \sum_k b_{j,k} = 1$$

Any of the state-dependent probability density functions (PDF) $b_j(x) = p(O_t = x \mid q_t = s_i)$ of an HMM with continuous output can be reasonably well approximated by a convex combination

$\sum_{k=1}^K b_{j,k} g_{j,k}(x) = \sum_{k=1}^K b_{j,k} N(x \mid \mu_{j,k}, \Sigma_{j,k})$ of multivariate Gaussian PDFs. The huge amount of statistical parameters found in a continuous mixture HMM as defined above – the model includes estimates of a distribution mean $\mu_{j,k} \in \mathbb{R}^d$ and a covariance matrix $\Sigma_{j,k} \in \mathbb{R}^{d,d}$, for each of $(N \cdot K)$ index pairs can be drastically reduced if all state-dependent sets $\{g_{j,k} \mid K=1, \dots, K\}$ of mixture components are pooled into one global collection of PDFs. The resulting type of model is termed semi-continuous HMM; its output distributions

$b_j(x) = \sum_{k=1}^K b_{j,k} g_k(x)$ with $g_k(x) = N(x \mid \mu_k, \Sigma_k)$ all share the same global set $g = \{g_1, \dots, g_K\}$ of Gaussians regardless of the state index j . The semi continuous model is therefore characterized by the statistics $\lambda = (\pi, A, B, g)$, where density function g_k is represented parametrically by its mean vector $k \mu$ and covariance matrix Σ_k . Evidently, our notation suggests that π, A , and B can be interpreted as an ordinary discrete HMM and g as the codebook of a K - class probabilistic (soft) vector quantizes, transforming continuous feature vectors x into a likelihood array $(g_1(x), \dots, g_K(x))^T$.

2.4 Support Vector Machines (MHSVM)

The utilization of support vector machine (SVM) [9], [10] classifiers has gained immense popularity in the last years. SVMs have achieved excellent recognition results in various pattern recognition applications [10]. Also in offline optical character recognition (OCR) they have been shown to be comparable or even superior to the standard techniques like Bayesian classifiers or multilayer perceptrons [11]. SVMs are discriminative classifiers based on Vapnik's structural risk minimization principle. They can implement flexible decision boundaries in high dimensional feature spaces. The implicit regularization of the classifier's complexity avoids over fitting and mostly this leads to good generalizations. Some further properties are commonly seen as reasons for the success of SVMs in real-world problems: the optimality

of the training result is guaranteed, fast training algorithms exist and little a-prior knowledge is required, i.e. only a labeled training set. Here, we provide a brief introduction to support vector classification.

For more details and geometrical interpretations please refer to the standard literature, e.g. by Burges [9] or Cristianini and Shawe-Taylor [10].

Consider a two-class classification problem and a set of training vectors $\{p_i\}_{i=1,\dots,M}$ with corresponding binary labels $S_i=1$ for the "positive" and $S_i=-1$ for the "negative" class. In classification an SVM assigns a label 'S' to a test vector T by evaluating

$$f(T) = \sum_i \alpha_i S_i K(T, P_i) + b \text{ and } S' = \text{sign}(f(T))$$

The weights α_i and the bias b are SVM parameters and adopted during training by maximizing

$$L_D = \sum_i \alpha_i - (1/2) \sum_{i,j} \alpha_i \alpha_j S_i S_j K(P_i P_j)$$

under the constraints $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i S_i = 0$

with C a positive constant weighting the influence of training errors. $K(\cdot, \cdot)$ is the kernel of the SVM. A solution for the α_i implies a value for b . The SVM framework gives some flexibility in designing an appropriate kernel for the underlying application. Many implementations of kernels have been proposed so far, one popular example is the Gaussian kernel

$$K(P_i, P_j) = \exp(-r \|p_i - p_j\|^2)$$

If $K(\cdot, \cdot)$ is positive definite, (1)-(2) is a convex quadratic optimization problem, for which the convergence towards the global optimum can be guaranteed. However, obtaining this solution for real-world problems can be quite demanding and requires sophisticated optimization algorithms like chunking, decomposition or sequential minimal optimization [10]. Usually $\alpha_i = 0$ for the majority of i and thus the summation in (3) is limited to a subset of the P_i , which therefore is called the set of support vectors. Extensions of the binary classification to the multi-class situation are suggested in several approaches [9], [12].

2.5 Radial Bass Function Neural Network (RBF-NN)

To improve the accuracy, we have trained RBFNN with the output of both the algorithms. Different samples of Tamil Characters are taken and given as input to both HMM and SVM. If HMM or SVM gives a false character, the neural network is trained with the weightage of both the algorithms and the actual character. This process is done for all the possible false recognition of the two algorithms. During OCR When both the algorithms not giving same character, trained RBFNN is used to retrieve the actual character. This way we can increase the accuracy of OCR. Radial Basis Functions emerged as a variant of artificial neural network in late 80's. RBF's are embedded in a two layer neural network, where each hidden unit implements a radial activated function [13].

Due to their nonlinear approximation properties, RBF networks are able to model complex mappings, which perception by means of multiple intermediary layers [14]. Radial basis networks can require more neurons than standard feed forward back propagation networks, but often they can be designed in a fraction of the time it takes to train standard feed forward networks. They work best when many training vectors are available. RBF networks have been successfully applied to a large diversity of applications including interpolation [15], image restoration [16], shape-from-shading [17], 3-D object modeling [18], data fusion [19], etc.

2.6 Encoded Character String Dictionary

Dictionaries [15] obviously look like texts and share many features with other types of texts. However, users typically do not read a dictionary linearly, but access entries on the basis of a *key* (the headword) in order to retrieve various fields of information associated with that key (pronunciation, grammatical information, etymology, definitions, etc.). Electronic dictionaries now commonly available on CD ROM. In addition, although the display on the screen still looks more or less like a text, the internal representation is rarely that of a linear text. Here, a list of tamil words can be stored which will be used to find the occurrence of next letter of a word. This will enable the Neural Network algorithm to guess the word with more efficiency in case of confusing letters like la and va. These encoded dictionaries can be integrated with the neural network for efficiency in recognizing the characters.

3. Conclusion

In this paper we have proposed a new method of Tamil OCR where we have fused three algorithms to get the maximum possible efficiency by integrating it with a encoded character string dictionary. Our work primarily deals with choosing a better algorithm, and fusing them finally to attain better accuracy. We have chosen HMM, SVM to be fused and finally used Neural network to predict the correct character when there arises a situation where two algorithms yield two different characters. Our future work is to increase the efficiency more by increasing the effectiveness of neural network.

4. References

- [1] G. Nagy. Twenty years of document image analysis in pattern analysis and machine intelligence. IEEE Tran. PAMI, pages 38-82, 2000.
- [2] Mantas, J., 1986. An overview of character recognition methodologies, Pattern recognition, 19 (6): 425-430.
- [3] Govindan, V.K. and A.P. Shivaprasad, 1990. Character Recognition-A Review, Pattern Recognition, 23 (7): 671- 683.
- [4] Hewavitharana, S, and H.C. Fernando, 2002. A Two Stage Classification Approach to Tamil Handwriting Recognition, pp: 118-124, Tamil Internet 2002, California, USA.
- [5] Chinnuswamy, P., and S.G. Krishnamoorthy, 1980. Recognition of Hand printed Tamil Characters, Pattern Recognition, 12: 141-152.
- [6] R.M. Suresh, S. Arumugam and K.P.Aravanan, "Recognition of handwritten Tamil characters using fuzzy classificatory approach ", Proc. The Tamil Internet 2000 Conference, Singapore, July 2000.
- [7] Siromoney et al., 1978. Computer Recognition of Printed Tamil Character, Pattern Recognition 10: 243-247.
- [8] H. Bunke, M. Roth, and E. G. Schukat-Talamazzini. Offline Cursive Handwriting Recognition using Hidden Markov Models. Pattern Recognition, 28(9):1399-1413, 1995.
- [9] C. Burges. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2):121-167, 1998.
- [10] N. Cristianini and J. Shawe-Taylor. Support Vector Machines. Cambridge University Press, 2000.
- [11] D. DeCoste and B. Schölkopf. Training invariant support vector machines. Machine Learning, 46(1/3):161, 2002.

- [12] G. C. Cawley. MATLAB support vector machine toolbox (v0.50 β). University of East Anglia, School of Information Systems, Norwich, Norfolk, U.K. NR4 7TJ, 2000.URL <http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox>.
- [13] Adrian G. Bors, "Introduction of the Radial Basis Function (RBF) Networks", Department of Computer Science, University of York, UK.
- [14] Haykin, S. (1994) Neural Networks: A comprehensive Foundation. Upper Saddle River
- [15] Nancy Ide and Jean Veronis, "Encoding Dictionaries", Department of Computer Science, Vassar College, Poughkeepsie, New York 12601 (U.S.A.).

Recognition of Ancient Tamil Characters in Stone Inscription

S.Rajakumar, Asst.Professor, Department of ECE

Dr.V.Subbiah Bharathi, Dean (A), Department of ECE

J. Navarajan, Professor, Department of ECE

DMI College of Engineering, Chennai-602103, India

E-mail:rajkumarmeae135@yahoo.co.in

Abstract

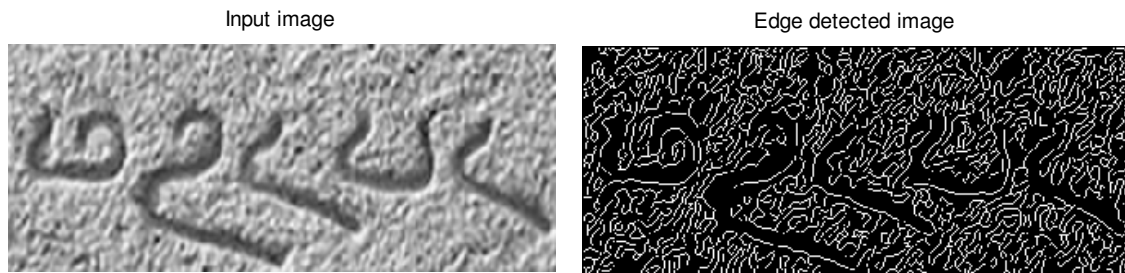
In our project, we simulate a technique for recognizing ancient Tamil characters in stone inscription of various centuries. Many medical mysteries and valuable historical secrets are hidden in stone inscription which only few people can understand. Though there are number of character recognition, there is no efficient system for Tamil language which has a large character set and huge amount of variation. In particular, there is no system that addresses either the morphological characteristics of stone inscriptions or accounts for ancient characters. We propose a 3-stage approach to increase the recognition efficiency of characters. The image of the sculpture is captured and converted to required dimensions. Segmentation of the character is done using Chan-veye algorithm which uses active contours for separating the characters from background. Chan-veye algorithm has resistance to the morphological characteristics of stone inscriptions as it uses active contours. The noise in the image is removed using the area property. Zoning of characters is done by which each character is divided into six regions each viz., three horizontal and three vertical regions. Feature extraction is carried out by extracting the geometric features of the character contour. These features are based on the basic line types that form the character skeleton. 54 features (nine for each region) representing the number and normalized length of line segments of is extracted. The feature vectors so generated are applied to neural network (which is trained in advance with the features extracted from the characters of training database) for pattern recognition. Post processing of the pattern obtained is done to identify the exact character.

Objective: The main objective of this thesis is to convert ancient Tamil characters of various centuries to modern character by training the computer. This thesis explains how the characters are extracted and recognized using neural networks.

Problem Definition: Though there are a number of Character Recognition systems available, Character recognition in Stone Inscription is still a challenging task due to the following factors:

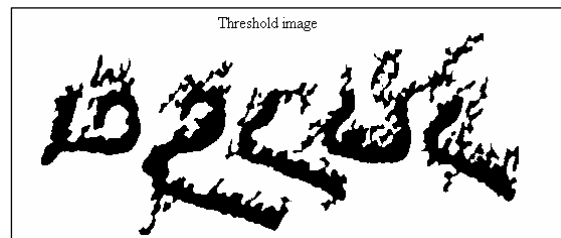
- Texture of inscriptions
- Noise present in image
- Complex characters
- Matching characters of different centuries
- Large character set
- Size of letters etc.

Existing Systems:



Edge detection is performed by computing gradients in an image. The contour of points with maximum gradient forms the edges in the image. Various operators used for these systems are Sobel operator, Roberts operators, Prewitt operators, Canny operator etc. These systems cannot be applied to stone inscriptions as the surface of the stones has ruggedness.

As shown in the above figure edge detection systems cannot be used in stone inscriptions due to its texture. Also the morphological characteristics of the stone inscriptions must be addressed during segmentation. In Thresholding, the mean values of the pixel values are calculated. Then, the histogram of the image is obtained. The mean value of the pixels is applied over the histogram to obtain the threshold value. Once the threshold value is obtained, all pixels having intensity greater than threshold are converted to the value of 255 and rest of the pixels is converted to 0. By performing the above operation, Binary image of the image is obtained having the value of 255 for characters and 0 for background of the image.



As shown in the above figure, when thresholding is applied the neighbours of the character is also taken as character due to the degradation function of the image. When the degradation function becomes significantly higher, the character cannot be segmented successfully.

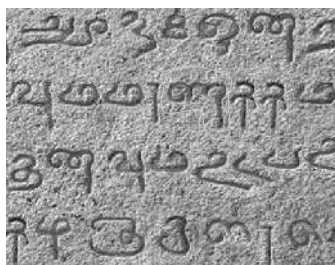
Proposed System :

In our project, we propose a morphological segmentation that utilises the features of adaptive thresholding combined with dilation of centroids that eliminate the disadvantages of classic thresholding methods. The various steps involved in it are

- Specifying the ROI
- Morphological segmentation
 - Conversion to Gray image
 - Adaptive thresholding
 - Applying Median filter
 - Use of Centroids

- Universe of Discourse
- Auto-correlation
- Neural networks

Specifying the ROI: Filtering a region of interest (ROI) is the process of applying a filter to a region in an image, where a binary mask defines the region. For example, you can apply an intensity adjustment filter to certain regions of an image. The multiROI function is used to obtain the region of interest from the image. The number of region of interest must be specified prior to running the program. Once it is specified, a figure window opens as shown in the image. The required region is selected by the polygon outline. The mask representing the region of interest is formed as shown in figure. This mask is used to extract the required character from the image.



ROI Region



ROI Mask

Morphological segmentation: A grayscale image (also called gray-scale, gray scale, or gray-level) is a data matrix whose values represent intensities within some range. MATLAB stores a grayscale image as a individual matrix, with each element of the matrix corresponding to one image pixel. In RGB format, each pixel has three values one each for Red, Green and Blue colours. The mean value of the image for each colour are different and hence thresholding cannot be applied directly for RGB image. Hence it is converted into gray image before applying thresholding. The Gray value of the pixel can be calculated as $I = .3(R) + .59(G) + .11(B)$

Where,

- I = Gray image intensity value
- R = Red component of RGB image
- B = Blue component of RGB image
- R = Red component of RGB image

Once the gray image is obtained, adaptive thresholding is done to convert it into a binary image. The histogram of the image is calculated for the image and the threshold value is calculated. All pixels having value greater than the threshold are converted into 0 and other pixels are converted into 1 as shown in figure. Median filtering is similar to using an averaging filter, in that each output pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel. However, with median filtering, the value of an output pixel is determined by the median of the neighborhood pixels, rather than the mean. The median is much less sensitive than the mean to extreme values (called outliers). Median

filtering is therefore better able to remove these outliers without reducing the sharpness of the image. The `medfilt2` function implements median filtering. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. The noise which occupies large pixel values cannot be removed using median filter. The area property of the bounded region is used to remove these noises. The `bwlabel` function is used for implementing the above step. The command, `L = bwlabel(BW,n)`, is used which returns a matrix `L`, of the same size as `BW`, containing labels for the connected objects in `BW`. `n` can have a value of either 4 or 8, where 4 specifies 4-connected objects and 8 specifies 8-connected objects; if the argument is omitted, it defaults to 8. The elements of `L` are integer values greater than or equal to 0. The pixels labeled 0 are the background. The pixels labeled 1 make up one object, the pixels labeled 2 make up a second object, and so on. The `regionprops` command is then applied over the matrix, `L` to obtain the area of each bounded region.

The threshold of area is selected as 500 and bounded regions having area lesser than this are eliminated. The resulting image contains only the binarised characters. The region of interest mask obtained from the `multiROI` function is overlapped with the noise removed image to extract the specified character from the image. Once the required character is segmented universe of discourse function is applied over it to obtain the boundary of the character. Due to this function, the variations in the size of letters is removed as all characters will be transformed into the same size.

Universe of discourse function finds the lowest number of the row that contains a non-zero element(a), the highest value of the row that contains non-zero element(b), the lowest number of the column that contains a non-zero element(c), the highest value of the column that contains non-zero element(b). These values are used to extract the required character. The required character can be now described as bounding box extending from `a` to `b` in x-direction and `c` to `d` in y-direction. The character obtained is complemented and converted to the dimension of 256 X 256 before comparing with the database to make it compatible with the images in the database.

The character after being resized is cross-correlated with every image in the database. In image processing, cross-correlation is a measure of similarity of two images as a function of a time-lag applied to one of them. This is also known as a sliding dot product or inner-product. The correlation is done in 2 dimensions and is obtained as

$$R(i, d) = \frac{E[(X_i - \mu_i)(X_d - \mu_d)]}{\sigma_i * \sigma_d}$$

Where

- 'i' is the input character
- 'd' is the image from database
- 'E' represents expectation value
- 'X' represents mean
- 'μ' represents intensity of individual pixel
- 'σ' represents the standard deviation

The above equation is applied over every pixel in each image in database. The correlated values obtained are given as input to the neural network. The neural network selects the image that has maximum correlation in the database as the recognised character.

OUTPUT IMAGE:



References:

1. An adaptive Technique for handwritten Tamil Character Recognition-IEEE 2007, Sarveswaran.K and Ratnaweera.D.A.A.
2. Enhancing the performance of handwritten Tamil character recognition by slant removal and introducing special features.-Journal of soft computing-2008.N.Shanthi and K.Duraiswamy.
3. Recognition of hand printed Tamil Characters, Pattern Recognition -1980, Chinaswamy and S.G.Krishnamoorthy.

Design and Evaluation of Omnifont Tamil OCR

Tushar Patnaik, CDAC Noida
tusharpatnaik@cdacnoida.in

Shalu Gupta, CDAC Noida
shalugupta@cdacnoida.in

CV Jawahar, IIIT Hyderabad
jawahar@iiit.ac.in

Santanu Choudhury, IIT Delhi
santanuc@ee.iitd.ernet.in

A G Ramakrishnan, IISc, Bangalore 560012.
ramkiag@ee.iisc.ernet.in

IISc Bangalore has developed a recognition engine for Tamil printed text, which has been tested on 1000 document images of pages scanned from books printed between 1950 and 2000. IIIT Hyderabad has developed a XML based annotated database for storing the 5000 images of scanned pages and the corresponding typed text in Unicode. CDAC, Noida has developed an efficient evaluation tool, which compares the OCR output text to the reference typed text (ground truth) and flashes the substitution, deletion and insertion errors in different colours on the screen, so that the design team can quickly identify the issues with the OCR and make corrective steps for improving the performance. IIT Delhi has proposed and developed a novel scheme for segmenting only the text regions from document images containing pictures. The OCRs uses Karhunen-Louve transform (KLT) as features and a support vector machine (SVM) classifier with RBF kernel in a discriminative directed acyclic graph (DDAG) configuration. They assume an uncompressed input image of the document page, scanned at a minimum of 300 dpi with 256 gray levels (not binary or two-level). Tamil OCR currently gives over 94% recognition accuracy at the Unicode level, evaluated on over 1000 printed pages, some of them also containing old Tamil letters.

The features of the OCRs are:

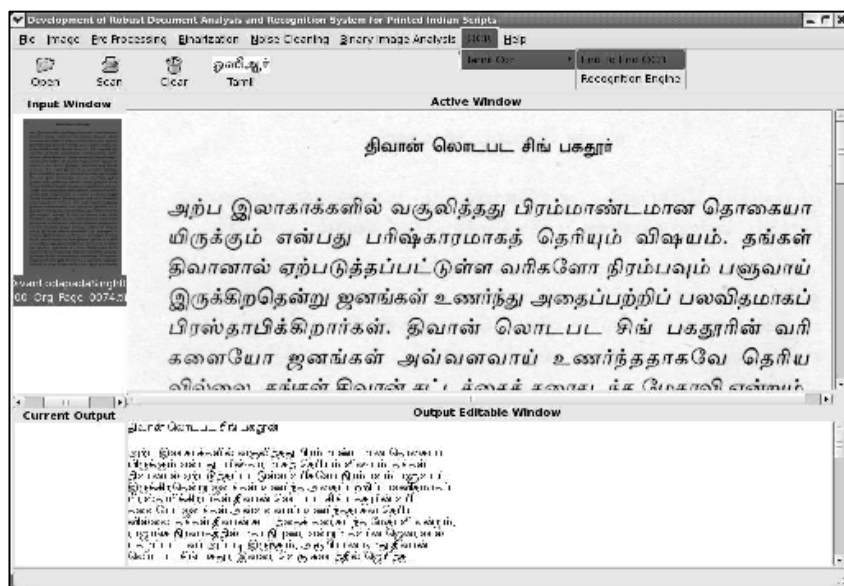
15. **Omnifont** : Any normal font used by books is handled. We don't say it is font-independent, because ornamental or stylized fonts cannot be handled.
16. **Merged Characters**: To a certain extent, the OCR is capable of identifying and segmenting the merger between two adjacent characters in a old, printed book.
17. **Noise Tolerance**: Certain types of breaks in the character are handled successfully.
18. **Old Tamil or Kannada Script**: The pre-1970 (prior to script revision due to E V Ramasamy Naicker) Tamil CV combinations such as NA, RA, nA, Nai, lai, Lai and nai are all recognized, along with the revised representations of the same. Similarly, old Kannada (*halegannada*) characters of *La* and *zha* and their vowel combinations are all handled seamlessly.

19. **Unicode Output:** The output is given in UTF-8.
20. **Testing:** Both OCRs have been tested by CDAC, Pune using an annotated corpus of over 1,000 document images of pages scanned from books printed between 1950 and 2002.
21. **Consistency:** The OCRs produce consistent and graded performance with font, size and quality variations.
22. **Future Enhancement:** The current average performance of 94% for Tamil and 84% for Kannada at the character level is without the use of any language model for postprocessing. Thus, there is a good potential to improve the performance of both OCRs further.

Medical Intelligence and Language Engineering Laboratory has teamed up with Bookshare.org, an International non-profit organization, to provide Tamil and Kannada digital books (copyright free or permitted by authors) online to print-disabled people (visually challenged, old people with vision disabilities and people with other disabilities that make it impossible for holding a book and turn pages of it). A Text-to-speech engine in the respective language will also be provided to the registered user, who can then directly listen to the printed content on their desktop or laptop. We look forward to partners, who can give us copyright free books (hard or soft copies) or direct us to sources of the same. They are also welcome to directly partner with bookshare.org or Worth Trust at Chennai or Enable India at Bangalore.

Figure 1 shows a screen shot showing the performance of the system, as well as the convenience and use of the GUI. Figure 2 shows the confusion matrix shown by the evaluation tool, which helps in identifying common confusion and improve the OCR accordingly. Figure 3 shows the evaluation tool, comparing the XML annotated Tamil text for the page and the OCR's output in Unicode. Use of such convenient tools accelerated the development of the OCR accuracy, and it is currently giving a performance of over 95% for good quality printed pages.

Fig.1. A screen shot from the OCR, showing the input image and the output text.



Annotation

திவான் லொடபட சிங் பகதூர்

அற்ப இலாகாக்களில் வசூலித்தது பிரம்மாண்டமான தோகையா யிருக்கம் என்பது பரிஷ்காரமாகத் தெரியும் விஷயம். தங்கள் திவானால் ஏற்படும் தப்பாட்டுள்ள வரிகளோ நிரம்பவும் பணுவாய் இருக்கிறதென்று ஜனங்கள் உணர்ந்த அதைப்பற்றிப் பலவிதமாகப் பிரஸ்தாபிக்கிறார்கள். திவான் லொடபட சிங் பகதூரின் வரி களையே ஜனங்கள் அவ்வளவாய் உணர்ந்ததாகவே தெரிய வில்லை. தங்கள் திவான் சட்டத்தைக் கரணகடந்த மதோவி என்றும், ராஜாங்க நிர்வாகத்தில் மகா நிபுணர் என்றும் கவர்னர் ஜெனரலால் புகழப்பட்டவர். அப்படி இராந்தம், அபியான நமது திவான் லொடபட சிங் பகதூர் இவரை வகை சூலயத்தில் ஜெயித்த, இவர் காரியத்திற்கு ஆகாது ஏட்டுச்சூரணைக்காய் என்று ராஜாபட்டத்தி விட்டார், இவர் எப்போதும் சட்டம், சாட்சியம் ஆகிய இரண்டையாமே மக்கியமாக எண்ணி, உயிரற்ற இயந்திரம் பரோல வலேன செய்கிறவரையென்றி, தம்மடைய பகத்தறிவன உபயோகப் பட்டி, யக்தா யக்தமறிந்த, ஜீவகாரணயம், பச்சாதாபம் மாதலிய கணங்களைக் கலந்துகொண்ட சட்டங்களை உபயோகப்பட்டத்த தெரிந்த கொள்ளாதவர், இவரக்க நல்ல பூத்தி புகடவலேண்டும் என்று நமது திவான் லொடபட சிங் பகதூர் அபியாக இராந்த இத்தனை காரியங்களையம் மகா அற்பாதமாக இயற்றிக் காட்டி இரக்கிறார் என்பது நன்றாகத் தெரிகிறது. ஓர் மனிதன் சாட்சிகள் இல்லாமலும், அல்லது, பறையச் சாட்சிகளை உண்டாக்கிக் கொண்டும், சட்டத்தில் அகப்பட்டிருக்கின்றோன்றாமல், எவ்வளவோ அபாரமான காரியங்களை எல்லாம் எத்தனை வருஷகாலம் வரையில் வலேண்டுமாளாவம் செய்வது சாத்தியமான விஷயம் என்பது இப்போது நன்றாக விளங்கிவிட்டது. ஓர்வளடைய சொத்தனை இன்னொருவன் அக்கிரமமாக அபகரித்தக் கொள்ள கிறான். ஆனால் அதற்கு சாட்சிகள் இல்லை. ஆகவே, சட்டத்தின் ப.ப.பி. பார்த்தகல் மஹ சங்கமகசிகரில்லலை சிமயப்படு பார்த்தகல்

OCR Output

திவான் லொடபட சிங் பகதூர்

அற்ப இலாகாக்களில் வசூலித்தது பிரம்மாண்டமான தோகையா யிருக்கம் என்பது பரிஷ்காரமாகத் தெரியும் விஷயம். தங்கள் திவானால் ஏற்படும் தப்பாட்டுள்ள வரிகளோ நிரம்பவும் பணுவாய் இருக்கிறதென்று ஜனங்கள் உணர்ந்த அதைப்பற்றிப் பலவிதமாகப் பிரஸ்தாபிக்கிறார்கள் திவான் லொடபட சிங் பகதூரின் வரி களையே ஜனங்கள் அவ்வளவாய் உணர்ந்ததாகவ தெரிய வில்லை. தங்கள் திவான் சட்டத்தைக் கரணகடந்த மதோவி என்றும், ராஜாங்க நிர்வாகத்தில் மகா நிபுணர் என்றும் கவர்னர் ஜெனரலால் புகழப்பட்டவர் அப்படி இராந்தம், அரலியான நமது திவான் லொடபட சிங் பகதூர் இவரை வகை சூலயத்தில் ஜெயித்த, இவர் காரியத்திற்கு ஆகாது ஏட்டுச்சூரணைக்காய் என்று ராஜாபட்டத்தி விட்டார் இவர் எப் போதும் சட்டம், சாட்சியம் ஆகிய இரண்டையாமே மக்கியமாக எண்ணி, உயிரற்ற இயந்திரம் பரோல வலேன செய்கிறவரையென்றி, தம்மடைய பகத்தறிவன உபயோகப்

0086_DivanLodapadaSinghBagadoor_Img_300_Org_Page_0074

AnnotationSize	OCRoutputSize	Substitution Error	insertScore	deleteScore	Substitution Error-Rate	Total Error-Rate
1569	1536	10	5	38	0.6373%	3.378%

Note:- Substitution, Insert, Delete

Previous Next Reset End

Fig. 3. Evaluation Tool showing substitution, insertion and deletion errors.

A Novel Hybrid SVM-Neural Approach to Recognize Handwritten Tamil Characters

N.Shanthi

E-mail:shanthimoorthi@yahoo.com

K.Duraiswamy

E-mail:drkduraiswamy@yahoo.co.in

K.S.Rangasamy College of Technology

Tiruchengode, Tamil Nadu, India

Abstract

In this paper a Tamil handwritten character recognition algorithm based on hybrid SVM-Neural approach is presented. Initially SVM and Neural network is individually tested to know the performance of each classifier. The recognition accuracy of SVM is 90% and the recognition accuracy of neural network is 81%. So a hybrid approach is used to improve the recognition accuracy.

Data samples are collected from different writers on A4 sized documents. They are scanned using a flat bed scanner at a resolution of 300dpi and stored as grey scale images. Various preprocessing operations are done on the input document image to enhance the quality. The characters are segmented and normalized to uniform size of 64 X 64. These uniform sized characters are projected onto a grid of fixed size 8 X 8. The pixel density is calculated for each grid and used as the feature vector. These features are given to well known support vector machines in the first stage. Few characters with low recognition accuracy are selected and are given to neural classifier. Since few classes are given to the neural network the recognition accuracy is improved. The new hybrid approach yields a recognition accuracy of 91.25%.

1. Introduction

Machine simulation of human reading is one of the areas, which has been the subject of intensive research for the last three decades, yet it is still far from the final frontier[1][2]. Works on offline recognition of handwritten characters has been done well on Chinese, Arabic and few other scripts of the other nations[13][14][15]. However, there is less progress towards recognition of handwritten characters of Indic scripts. Recognition of handwritten Indian scripts is difficult because of the presence of vowel modifiers and compound characters. Most of the existing works are concerned about Devnagari and Bangla script characters, the two most popular languages in India. Few works are reported on the recognition of other languages like Telugu, Oriya, Kannada, Punjabi, Gujarathi [4] etc., There are few attempts to recognize printed or handwritten Tamil characters [3][6][8][9] [17][18]. This paper proposes a recognition system for handwritten Tamil characters. The paper is organized as follows. The following section presents a brief description of Tamil language. The proposed hybrid SVM-Neural approach is presented in section 3. Finally, in section 4 conclusion and future works are given.

2. Tamil Language

Tamil, which is a south Indian language, is one of the oldest languages in the world. Sanskrit has influenced it to a certain degree. Its history can be traced back to the age of Tolkappiyam, the earliest extant Tamil grammar generally to 500 B.C. Tamil language is a member of the Dravidian / South Indian

family of languages. Among the Dravidian language it is least influenced by 'Sanskrit' though there is a certain degree of influence. Most Tamil letters have circular shapes partially due to the fact that they were originally carved with needles on palm leaves, a technology that favored round shapes. The Tamil script is used to write the Tamil language in TamilNadu, SriLanka, Singapore and parts of Malaysia.

Tamil is the official language of the Indian state of TamilNadu, classical language in India, and has official status in India, Sri Lanka and Singapore. With more than 77 million speakers, Tamil is one of the more widely spoken languages in the world. The script for Tamil Language was conceived since time immemorial and the present form is after undergoing various changes at various periods of time. During its long cherished existence, it underwent various changes, corrections, modifications and has arrived at the present form based on its applicability and adoptability in the ever changing writing method and devices from time to time. The present form of the Tamil script is believed to have been evolved based on its "adoptability" for handwriting mechanisms.

2.1 Tamil Alphabets

Tamil alphabet consists of 12 vowels, 18 consonants, a special character (akh) and 216 combinational alphabets obtained by combining the vowels and consonants. Tamil language is relatively simpler compared to other Indian scripts. The rules for character composition are far fewer than in other Indian scripts. The only category of composition allowed is of Consonant-Vowel type, where a structure corresponding to a consonant and another corresponding to vowel are combined to form a C-V type character with a unique shape. Even though the total number of alphabets is 247, the graphical characters required to represent all of them are only 106. The C-V combination results in unique shape when அ, இ, ஈ, உ, ஊ are combined with consonants. When the remaining vowels are combined with consonants, ஊ then the combination results in horizontally isolatable structure. For example by adding ஃ with அ we get ஊ . Similarly when all the consonants are added with அ , we get 18 unique characters. Then by placing the character ஊ after these character we get one alphabet. Following ஊ we get ஊஈ ; following ஊ we get ஊஊ ; following ஊ we get ஊஊ . In this manner there are eighteen alphabets with the character ஊ coming after different consonants. So the composite characters can be recognized as the sequence of ஊ, ஊ, ஊ characters. Due to this flexibility the total number of characters to be recognized is reduced to 106. If these 106 characters are recognized then all the 247 characters can be recognized. In addition to this the other unique characters to be recognized are ஊஊ , and ஊ .

3. Hybrid SVM-Neural approach

For relatively small class problems a single classifier system provides acceptable accuracies. But when the number of classes is more, the recognition accuracy is less. So hybrid approach can be used to recognize complex character recognition problem.

Classification with multiple classifiers has been an active research area since early 1990s. Many combination methods have been proposed, and the applications to character recognition problems have advantage over individual classifiers. Generally, a character recognizer involves three tasks: preprocessing, feature extraction and classification.

3.1. Preprocessing

Various preprocessing operations are performed prior to recognition to enhance the quality of the input image.

Preprocessing operations performed prior to recognition are:

1. **Thresholding** is to convert a gray scale image into binary by determining a value for gray-scale (or threshold) below which the pixel can be considered to belong to the writing and above which to the background; Here Otsu's method of histogram-based global thresholding algorithm is used[11].
2. **Skew detection and correction** is to detect the skew present in the image and to correct it. A document skew detection algorithm using wavelet transform and horizontal projection profile [19] is used in this work.
3. **Line segmentation**, the separation of individual lines of text; Horizontal histogram profile is used for segmenting the lines[10].
4. **Slant correction**, Removal of angle between the vertical direction and the direction of the strokes; a new method for slant estimation is used based on the combination of the projection profile technique and wavelet transform.
5. **Character segmentation**, the isolation of individual characters; Vertical histogram profile is used for segmenting the characters[10].
6. **Skeletonization** is the process of peeling off of a pattern as many pixels as possible without affecting the general shape of the pattern [12]; Here Hilditch's algorithm is used for skeletonization.
7. **Character size Normalization**, Size normalization is a transformation of an input image of arbitrary size into an output image of a fixed pre-specified size, while attempting to preserve structural detail[7]; Bilinear interpolation technique is used to convert the random sized image into normalized image of size 64X64.

Figure 1 shows the sample input image and Figure 2 shows the preprocessed image of the last line.

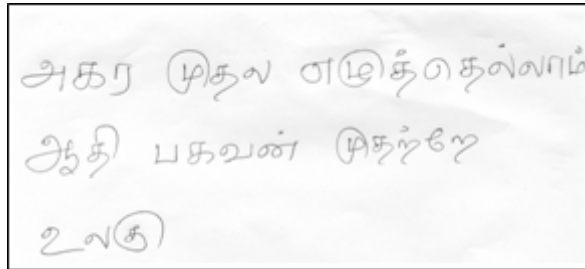


Figure 1. Sample input image

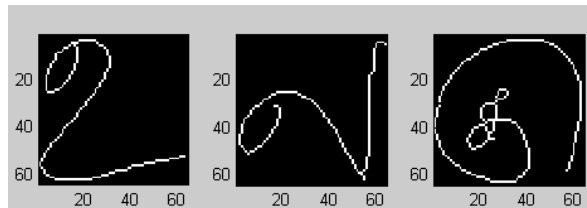


Figure 2. Preprocessed image

3.2. Feature extraction

Feature extraction is the problem of extracting the information from the preprocessed data, which is most relevant for classification purposes, in the sense of minimizing the within-class pattern variability, while enhancing the between-class pattern variability [20]. Selection of feature extraction method is the most important factor in achieving high recognition performance [5]. Here a simple feature of pixel density is used as feature. An 8 X 8 grid is superimposed on the character image, and for each of the 64 zones the average number of pixels is computed which results in 64 features. The pixel density varies from 0 to 64. The Figure 3 shows the zoning of character image.

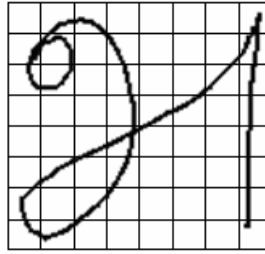


Figure 3. Zoning of character image

3.3. Classification

The final goal of character recognition is to obtain the class codes of character patterns. On segmenting character or words from document images, the task of recognition becomes assigning each character or word to a class out of a pre-defined class set. Different classifiers are used for recognition. Support vector machine and artificial neural network are used in this work for classification.

3.3.1 Support vector machines

The SVM is a new type of hyperplane classifier, developed based on the statistical learning theory of Vapnik[16], with the aim of maximizing a geometric margin of hyperplane, which is related to the error bound of generalization. The application of SVMs to character recognition has yielded state-of-the-art performance. Initially SVM alone is used for training and classification of the Tamil characters.

SVMs are mostly considered for binary classification. For multiclass classification, multiple binary SVMs, each separating two classes or two subsets of classes, are combined to give the multiclass decision. Here “one-against-one” approach in which $k(k-1)/2$ classifiers are constructed and each one trains data from two different classes. For training data from the i^{th} and the j^{th} classes, the following binary classification problem is solved:

$$\min_{w^{ij}, b^{ij}, \xi_t^{ij}} \frac{1}{2} (w^{ij})^T w^{ij} + C \left(\sum_t (\xi_t^{ij}) \right)$$

subject to

$$\begin{aligned} ((w^{ij})^T \phi(x_t) + b^{ij}) &\geq 1 - \xi_t^{ij}, \text{ if } x_t \text{ in the } i^{\text{th}} \text{ class,} \\ ((w^{ij})^T \phi(x_t) + b^{ij}) &\geq -1 + \xi_t^{ij}, \text{ if } x_t \text{ in the } j^{\text{th}} \text{ class,} \\ \xi_t^{ij} &\geq 0. \end{aligned}$$

Here max-wins voting strategy is used for classification. The recognition accuracy using SVM is 90%.

3.3.2 Artificial neural networks

An artificial neural network is defined as a computing structure consisting of a massively parallel interconnection of adaptive neural processors. The main advantage of artificial neural networks lies in the ability to be trained automatically from examples and good performance with noisy data. Artificial neural networks have been widely used in the field of character recognition produced promising results.

Back propagation feed forward neural network (BPN) is used for training and testing the network. The recognition accuracy is only 81%. Neural network took large time to train and the recognition accuracy is also low. So we have decided to use the benefits of both the classifiers.

3.4. Introduction to SVM-Neural Classifier

Even though the recognition accuracy of SVM is higher when compared to the other classifiers, still there is a scope for further improvement. The analysis of the Tamil character set reveals that a few characters are closely identical to each other in structure. So when all the characters are given to SVM then there is a difficulty in recognizing the identical characters. The characters identified and their percentage of recognition using SVM is shown in table 1. For example when the classification result of character ன is analyzed, the result shows that many a times the character ன is recognized as னை or னா. This is because of the reason that these characters look identical in structure. Hence in order to solve this difficulty a hybrid two stage classifier approach is proposed.

Table 1. Characters with low recognition accuracy

S.No.	Characters with low recognition accuracy	Similar characters	Recognition accuracy %
1	ஏ	ஏ,ர	78.45
2	ஓ	ஓ	72.8
3	ஞ	ஞ	77.7
4	ழ	ழ	81.29
5	ன	னை,னா	74.85

The hybrid approach consists of 2 stages. SVM is chosen as the classifier in the first stage since it produced the highest recognition accuracy. When the performance of the SVM classifier is analyzed it illustrated that the 64 × 64 sized image with overlapping zones produced the best recognition accuracy. So those 225 features are calculated and given as input to the SVM classifier in the first stage. The output of SVM revealed that the recognition accuracy of the characters ஏ, ஓ, ஞ, ழ and ன are low because they are identical to other characters in structure. The groups of such identical characters are (ஏ,ஏ,ர), (ஓ,ஓ), (ஞ,ஞ), (ழ,ழ) and (ன,னா, னை).

In the second stage a separate classifier is designed for each group of identical characters. Both SVM and ANN are used in the second stage and the performances of both the classifiers are analyzed for all the groups. Table 2 shows the characters with low recognition accuracy, their identical characters and improved recognition accuracy using SVM-SVM approach. Table 3 shows the improved recognition accuracy of SVM-Neural approach. From the Tables it is evident that the performance of ANN is good in the second stage when compared to SVM. Hence in the proposed hybrid approach SVM and ANN are combined sequentially to improve the classification performance. This hybrid approach improves the overall recognition accuracy by 1.5%.

Table 2 Recognition Accuracy of SVM-SVM Approach

S.No.	Characters with low recognition accuracy	Identical characters	Recognition accuracy %	Improved recognition accuracy %
1	எ	எ,ர	78.45	85.08
2	ஓ	ஓ	72.83	80.93
3	ஞ	ஞ	77.71	85.71
4	ழ	ழ	81.29	87.71
5	ஊ	ஊ,ள	74.86	80.00

Table 3 Recognition Accuracy of SVM-Neural Approach

S.No.	Characters with low recognition accuracy	Identical characters	Recognition accuracy %	Improved recognition accuracy %
1	எ	எ,ர	78.45	90.61
2	ஓ	ஓ	72.83	84.39
3	ஞ	ஞ	77.71	89.71
4	ழ	ழ	81.29	89.47
5	ஊ	ஊ,ள	74.86	81.14

This hybrid approach improves the overall recognition accuracy by 1.5% from the original 90% and improves the recognition accuracy of each identical character by 6% to 12%. SVM and ANN are combined effectively in the proposed hybrid approach to improve the overall recognition accuracy and the recognition accuracy of identical characters.

The architecture of the proposed hybrid SVM - Neural approach is shown in Fig.4.

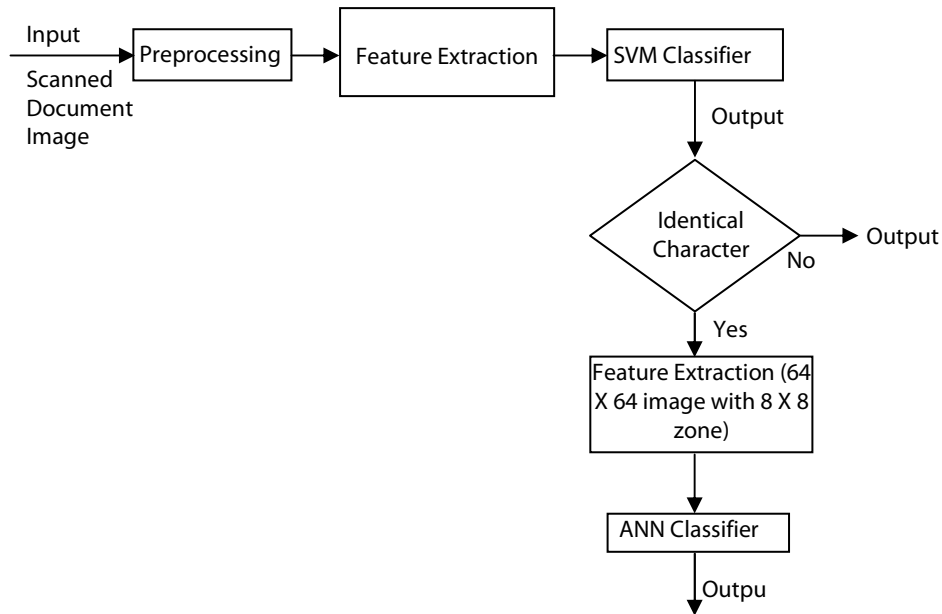


Fig.4 Architecture of the Proposed Hybrid SVM-Neural Approach

4. Conclusion

This study presents a novel hybrid system to recognize offline handwritten Tamil characters. The experimentation result shows that the simple SVM approach produces a cumulative recognition accuracy of 90% and the recognition accuracy for different characters varies from 72% to 99%. When neural network is used for recognition, the recognition accuracy varies from 63% to 98% with an overall recognition accuracy of 81%. But when SVM-neural approach is used the recognition accuracy is increased to 91.5%. This recognition accuracy is achieved with a simple feature of pixel densities in various zones of the images. The main recognition errors were due to abnormal writing and ambiguity among similar shaped characters. Future work can include more robust invariant features to achieve better discrimination power and neural classifier can be constructed for other characters.

References

- 1) J.Mantas. An overview of character recognition methodologies. *Pattern recognition*, 19(6):425-430, 1986.
- 2) V.K.Govindan and A.P.Shivaprasad. Character Recognition - A Review. *Pattern Recognition*, 23(7):671-683, 1990.
- 3) P.Chinnuswamy and S.G.Krishnamoorthy. Recognition of Hand printed Tamil Characters. *Pattern Recognition*, 12(3):141-152, 1980.
- 4) U.Pal and B.B.Chaudhuri. Indian Script Character Recognition: a Survey. *Pattern Recognition*, 37(9):1887-1899, 2004.
- 5) Trier, Anil K. Jain and Torfinn Taxt. Feature Extraction Methods for Character Recognition - A Survey. *Pattern Recognition*, 29(4): 641-662, 1996.
- 6) G.Siromoney, R.Chandrasekaran and M.Chandrasekaran. Computer recognition of printed Tamil characters. *Pattern Recognition*, 10(4):243-247, 1978.
- 7) Srihari et al. On-line and Off-line Handwriting Recognition: A Comprehensive Survey. *IEEE PAMI*, 22(1): 63-84, 2000.
- 8) R.M.Suresh et al., Recognition of Hand printed Tamil Characters Using Classification Approach. *ICAPRDT'99*, 63-84, 1999.
- 9) S.Hewavitharana and H.C.Fernando, A Two Stage Classification Approach to Tamil Handwriting Recognition, *Tamil Internet 2002*, California, USA, 118-124, 2002.
- 10) N.Shanthi and Dr.K.Duraiswamy. Preprocessing algorithms for the recognition of Tamil Handwritten Characters. 77-82, *Third International CALIBER 2005*, Kochi.
- 11) N.Otsu. A Threshold Selection Method from Grey Level Histogram, *IEEE Trans. System Man and Cyber*. 9(1):62-66, 1979.
- 12) Lam, Lee, Suen. Thinning Methodologies - A Comprehensive Survey, *IEEE PAMI*, 14(9): 869-885, 1992.
- 13) S. N. Srihari, X. Yang and G. R. Ball. Offline Chinese Handwriting Recognition: An Assessment of Current Technology. *Frontiers of Computer Science in China*, 1(2):137-155, May 2007.
- 14) L.M.Lorigo, V.Govindaraju. Offline Arabic handwriting recognition: a survey, *IEEE PAMI*, 28(5):712 - 724, 2006.
- 15) S. Jaeger, M. Nakagawa, C.L.Liu. A Brief Survey on the State of the Art in On-Line Handwriting Recognition for Japanese and Western Script, *IEICE Conference on Pattern Recognition and Media Understanding*, 1-8, March 2002.

- 16) Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. 2003.
- 17) U.Bhattacharya, S.K.Ghosh, and S.Parui. A Two Stage Recognition Scheme for Handwritten Tamil Characters. 1: 511-515, ICDAR 2007.
- 18) N.Shanthi and Dr.K.Duraiswamy ,A novel SVM-based handwritten Tamil character recognition system. Pattern Analysis and Applications: Volume 13, Issue 2 (2010), Page 173
- 19) Shutao Li, Qinghua Shen, Jun Sun, "Skew Detection Using Wavelet Decomposition and Projection Profile Analysis", Pattern Recognition Letters 28, pp.555 - 562, 2007.
- 20) Trier, Anil K. Jain, Torfinn Taxt, "Feature Extraction Methods for Character Recognition - A Survey", Pattern Recognition, Vol.29, No.4, pp. 641 - 662, 1996.

Effective Tamil Character Recognition in Tablet PCs using Pattern Recognition

Ferdin Joe J
Department of Computer Science
& Engineering,
Einstein College of Engineering,
Tirunelveli, e-mail:
ferdinjoe@gmail.com

Dr. T. Ravi
Department of Computer Science
& Engineering,
KCG College of Technology,
Chennai. e-mail:
travi675@yahoo.com

Prof. R. Velayutham
Department of Computer Science
& Engineering,
Einstein College of Engineering,
Tirunelveli. e-mail:
rsvel_kumar@yahoo.co.uk

Abstract

Tamil Character Recognition is done for many applications using various novel methodologies. But the recognition of Tamil characters in Tablet PCs is a challenging one. Various methodologies like SVM, Preprocessing, interpolation etc were used in the past but were not satisfactory to friendly-user point of view. In this paper we introduce a new methodology pattern matching for Tamil Character Recognition. This method was tested to be a better approach for Gujarati scripts. As our Classical Language Tamil characters also has more curls and twists like Gujarati scripts, we utilized the same methodology for Tamil Character Recognition. This methodology uses a series of operations on the digitally written character. Scanning, Inversion, Grayscale conversion and trimming are done in the preprocessing section. Back propagation using neural networks in the middle tier and at last pattern matching algorithms are employed. In this last phase instead of pattern matching done, we come forward with a modified pattern matching algorithm. We have termed this algorithm as Thellodai Line Fitting Algorithm. As a result of this series of operations, the recognized output is obtained. For tablet PCs the output was simulated in Matlab's Neural Network and Image Acquisition toolbox. After analyzing the simulated results, the recognition rate of this methodology is found to be about 70 - 80% on an average where as the other methodologies were tantalizing between 60s and 70s for character recognition of the Classical Language Tamil.

Index Terms—Pattern Classification, Thellodai Algorithm

Introduction

This paper deals with the new methodology implied for the effective pattern matching of Tamil character recognition in tablet PCs. Tablet PCs English characters are easily recognized by just scribing on the touch module. But for Tamil Characters, the mechanism is not available as of now in the case of Tablet PCs. English characters are easy and faster to recognize because, it has only 26 characters. There lies a difficulty for Tamil Character Recognition. It has 247 letters excluding the old scripting letters. So there is a need to classify the Tamil characters in a simple way and reduce the character set to a minimal extent.

Gujarati scripts as in [1] are given with a set of matching. They are checked for templates and pattern matching of the particular character. This is possible for Gujrati scripts because the number of characters used are less. But when this technique is implied Tamil, certain difficulties are experienced in the

performance of execution. Tamil Characters are totally 247 in number, excluding the other language influential characters.

Similarly some methodologies are given for handwritten Tamil character recognition also. In section II, the difficulties in execution of the various methodologies are discussed. In section III, the newly proposed Thellodai line fitting algorithm is discussed in detail. In section IV, the proposed model is elucidated. In section V, the performance measures and the modules for real time implementation are discussed and finally section VI concludes the entire work with the possibilities to execute this project in real time.

Related Work

The base work is selected as [1], because Gujrati also has curvy writings similar to Tamil. So we are taking the technology used in [1] as the back bone for our proposed model. The model stated in [1] fails for Tamil because, Tamil characters are large in number and they take larger time when compared to the Gujrati scripts.

Various methodologies are being proposed for Tamil Handwritten Character Recognition. The models stated in [3], [5], [6], [7], has various good features for larger devices or enterprise level solutions. But these methodologies are found to be less effective when we take in the case for Tablet PCs.

Tablet PCs normally deal with the device driver synchronization, character compression and well classified set of characters. Tamil character set consists of 247 characters totally. In the previous models proposed in [3], [6] and [7], the classification is not efficient to the expected levels of Tablet PC synchronization. English characters find easier in the tablet PC because, it has only 26 singlet characters for classification. After thorough analysis on the Tablet PC architecture, the problem statement is framed as follows: "There lies a need for classifying all the 247 letters in Tamil to a maximal extent. The number of patterns used should be reduced to a minimal extent. A new algorithm based on line fitting for pattern recognition needs to be framed. This system should be developed in such a way that the device driver synchronization for Tablet PCs is to achieved."

Thellodai Algorithm

As mentioned in the problem statement at the end of the previous section, we have developed a new algorithm based on line fitting. We term it as Thellodai Line Fitting algorithm. Thellodai pronounced /**Thellodai**/. Thellodai means "Clear stream" in Tamil. As per this algorithm concerned, this has the well classified set of Tamil characters. All the 247 characters are classified efficiently and 35 patterns are framed. Many letters are recognized as a combination of two or more other patterns.

The Thellodai algorithm has a predefined trained set of 35 characters. These characters are termed as Frequently Used Patterns Uf, Rare Patterns Rp and Original Pattern Op.

Algorithm 1: Thellodai Line Fitting Algorithm

```
Function match()
{
  Var Uf, Rp, Op;
  Wait for input;
  Set the combination;
  Decide on the permutation;
  If permutation is Uf
  {
```

```

    The sub patterns are matched;
  }
  If permutation is Rp
  {
    Rp sets are matched;
  }
  If permutation is Op
  {
    Move to next input;
  }
  If match is formed
  {
    Indent to next space and wait for input
  }
}

```

The above algorithm is framed specially for the Tamil character set. The patterns possible are reduced to a minimal set of just 35 patterns. These patterns are studied for the usage. According to this study, many of the characters are not used nowadays for casual usage.

So we don't need to concentrate much on these characters. Thellodai algorithm is included as an additional feature in our proposed system and it will be good for rarely used characters also. The frequently used characters are indexed in an ajax based system and it will show the possibilities of character identification.

Proposed Model

The proposed model is based on the architecture mentioned in [1]. The system is simulated in matlab 7.0 using the image acquisition toolbox and neural network toolbox. The image acquisition toolbox is used for simulating the patterns and neural network tool box for simulating the frequently used pattern set.

Fig 1: Expected output for the Tamil Character Recognition in Tablet PCs (Modelled).

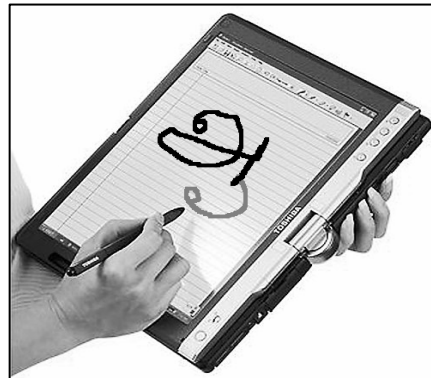
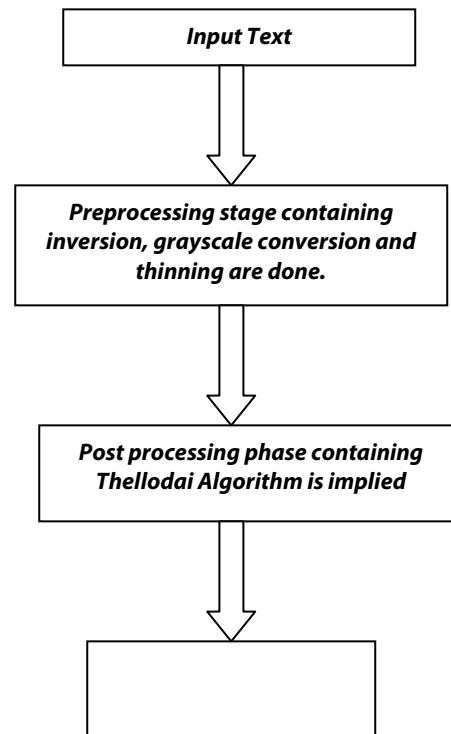


Fig 1 shows the expected output of the simulated set of data. For example, if we scribe "ah", it will give the possibility for "aah" also. The main aim is to study the efficiency of handwritten character recognition for aharam, aaharam, eharam, eeharam, uharam, ooharam, aeharam, aeharam, oharam, ooharam and ouharam set of characters. In Fig1, the character "ah" is scribed by the user. The system identifies as "ah" and gives the possible extension for "aah" also. The scribed character is displayed in black and the permuted extension is displayed in grey.

The recognition sector deals with the following processes: Scanning, Inversion, Grayscale conversion and trimming are done in the preprocessing section. Back propagation using neural networks in the middle tier and at last pattern matching algorithms are employed. In this last phase instead of pattern matching done, we come forward with a modified pattern matching algorithm. We have termed this algorithm as Thellodai Line Fitting Algorithm. As a result of this series of operations, the recognized output is obtained.



Preprocessing Phase

The preprocessing phase has the steps of Inversion, grayscale and thinning. During inversion, the character input is converted to the inverse color. Then during the grayscale phase, it gets converted to black and white. During thinning the edges are fitted.

Supervised phase

During the supervised phase, the Neural network principles come into act. The back propagation and trained intelligence modules will analyze the output of the preprocessing phase. Then by using the trained intelligence, the extensions are predicted. A popular and simple NN approach to the problem is based on feed forward neural networks with back propagation learning. In the training step, we each training sample is represented by two components: possible input and the desired network's output given that input. After the training step is done, we can give an arbitrary input to the network and the network will form an output, from which we can resolve a pattern type presented to the network. As per the requirement of the Neural Network algorithm the information content that is area within the borders needs to be digitized. This is achieved by dividing this area into a 12 x 12 grid. This 12 x 12 grid is scanned, and every time 1 is detected that particular block is flagged. This procedure is referred to as digitizing.

Fig2: Undigitized output for "ka"



The undigitized output is obtained in this phase. This output is now fed as input to the post processing phase.

0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	0
0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	0	0	0	0	1	0	1	0	0
0	0	1	0	0	0	1	1	0	1	0	0
0	0	1	1	0	1	1	0	1	1	0	0
0	0	0	1	1	1	0	1	1	0	0	0

Fig3: Digitized output as a result of the post processing phase

Postprocessing phase

Post processing phase is the final phase for this proposed methodology. This phase implies the Thellodai algorithm for the efficient pattern matching of the digitized tamil characters.

0	0	0	0	0	0	0	1	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	0
0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	0	0	0	0	1	0	1	0	0
0	0	1	0	0	0	1	1	0	1	0	0
0	0	1	1	0	1	1	0	1	1	0	0
0	0	0	1	1	1	0	1	1	0	0	0

Fig4: Output obtained as a result of Thellodai algorithm in the post processing phase.

The purple colored cell in Fig 4 indicates that the extension is possible for this character for “key” and “key” characters. When the user scribes on the purple extensions, it will automatically guide the user to incorporate the extension for the particular character to be finalized.

Performance Measures

The performance of the proposed methodology are found for the various subsets aharam, aaharam, eharam, eeharam, uharam, ouharam, aeharam, aeaharam, aiharam, oharam, ooharam and ouuharam.

The combination of 35 singlets is combined with each other and the characters are found for efficiency of matching. It is found that, the subsets of aharam, aaharam, aeharam, aeaharam, aiharam, oharam, ooharam and ouharam are in no need for any difficulty in recognizing the character. But eharam, eeharam, uharam, ouharam finds little difficulty. This is just for the reason for some characters like “Do” are in need of totally different shape transformation when compared with the other characters.

Character Subset	Recognition rate in %
Aharam	80
Aaharam	77
Uyir	79
Mei	80
Eharam	72
Eeharam	70
Uharam	71
Ouharam	69
Aeharam	76
Aeaharam	78
Aiharam	80
Oharam	75
Owharam	74
Ouharam	75

Table 1: Recognition rates of various subsets of Tamil character series.

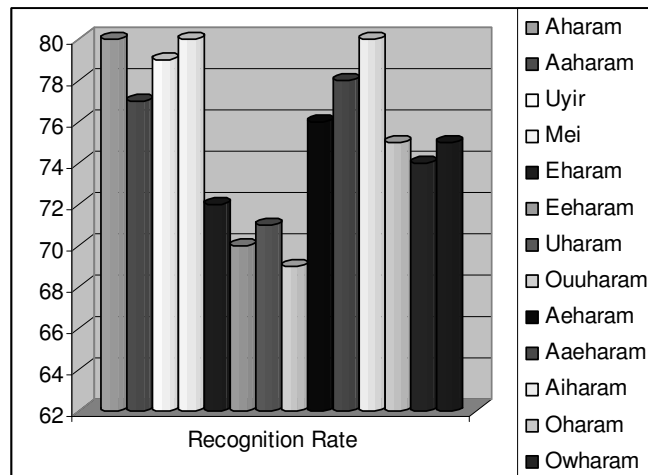


Fig 5: Graph showing the performance and recognition rates of various subsets of Tamil character set.

As shown in Fig 5, it is evident that the efficiency of the character recognition is found to be nearing 80% for most of the character subsets. Only 4 out of the 14 subsets, the efficiency is in the early 70s.

The recognition rates for eharam, eeharam, uharam, ouuharam are found to be nearing the early 70s because there are chances for asymmetric matching. When comparing with the recently existing systems mentioned in [1], [3], [13], it is clear that our proposed model will be better in its every aspect. The fidelity of the letters will be more. Handwritten performance of the user will increase.

There are many advantages for this system. When comparing to the transliteration system, the user is in no need to have English phonetic proficiency to type in Tamil. The proper pronunciations and spellings are possible. There is an issue of confusion with the lla and zha characters. This problem can be eliminated by using the proposed approach.

Tablet PCs as of now are not designed for Tamil Characters. This is because of the lack of proper hardware modifications. The modifications need to be done in such a way that Tamil characters can also be recognized. After this issue is fixed, real time implementation is possible for this proposed methodology.

Conclusion

From the above elucidated proposed methodology, it is clear that this technology is successful at the level of simulation. The efficiency for most of the cases comes above 75% and the Thellodai Line Fitting algorithm holds better than many of the existing pattern matching algorithms. The real time implementation needs the Unicode data of Tamil characters to be included in the Tablet PCs. If it is done so, then the real time implementation can be done. Tamil Tablet PCs are no more longer to wait.

References

- [1] Prasad JR, Kulkarni UV, Prasad RS, *Offline Handwritten Character Recognition of Gujrati Script Using Pattern Matching*. IEEE ASID 2009.
- [2] <http://www.touchscreens.com>
- [3] J. Sutha, N. Ramaraj, *Neural Network Based Offline Tamil Handwritten Character Recognition System*. IEEE Intl Conf on Computational Intelligence and Multimedia Applications 2007.
- [4] Ramanathan et al, *Tamil Font Recognition Using Gabor Filters and Support Vector Machines*, IEEE Intl Conf on Computing, Control and Telecommunication Technologies 2009.
- [5] KH Aparna et al, *Online Handwriting Recognition for Tamil*. Proc IEEE IWFHR-9 2004.
- [6] RM Suresh, L Ganesan, *Recognition of Printed and Handwritten Tamil Characters Using Fuzzy Approach*. Proc IEEE ICCIMA 05.
- [7] U Bhattacharya et al, *A two stage recognition scheme for Handwritten Tamil Characters*. Proc IEEE ICDAR 2007.
- [8] Sarveswaran K, Ratnaweera, *An adaptive technique for Tamil Handwritten Character Recognition*. Proc IEEE Intl Conf on Intelligent and Advanced Systems 2007.
- [9] Suresh Sundaram, AG Ramakrishnan, *A novel hierarchical classification scheme for online Tamil Character Recognition*. Proc IEEE ICDAR 2007.
- [10] Seethalakshmi et al, *Optical Character Recognition for printed Tamil text using Unicode*. Journal of Zhejiang University SCIENCE, 2005.

- [11] J. Venkatesh, C. Suresh Kumar, *Handwritten Tamil Character Recognition Using SVM*. Intl Journal of Computer and Network Security (IJCNS) Dec 2009.
- [12] Hiroki Mori et al, *Japanese Document Recognition Based on Interpolated n-gram model of character*. Proc IEEE ICDAR 1995.
- [13] R. Jagadeesh Kannan, R. Prabhakar. *An Improved Tamil Character Recognition System Using Octal Graph*, Journal of Computer Science 2008 Science Publication.

Ferdin Joe J is currently a PG scholar with the Department of Computer Science & Engineering, Einstein College of Engineering, Anna University @ Tirunelveli. He graduated in Computer Science and Engineering from Anna University @ Chennai during 2009. He is a member of IAENG and Chennai IT Professional User Group. He is an independent consultant and providing IT solutions for Academicians and Researchers. His areas of interest include Data Mining, Bioinformatics, SOA, Web 2.0 and Network Security. He has published nearly 10 papers in International Conferences, 1 paper in International Journal and 5 papers in National Conferences. He has a passion towards dedicated research in core computing principles. To know more about him and his research work, please log on to <http://www.ferdin.co.nr>.



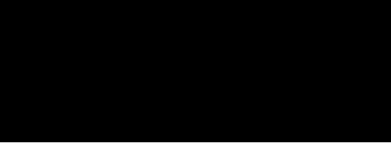
Dr. T. Ravi is currently Professor and Head with the Department of Computer Science & Engineering, KCG College of Technology, Anna University @ Chennai. He was previously with the Department of Computer Science & Engineering, Dr. MGR University, Chennai. He graduated in Computer Science & Engineering from Madurai Kamaraj University during 1991. He obtained Masters Degree during 1996 and PhD during 2009 in Computer Science & Engineering from Jadavpur University, Kolkata. He has nearly 18 years of teaching experience. He has published many papers in International Conferences and Referred International Journals. His areas of Interest include Data Mining, Networks and Bioinformatics. He is a recognized supervisor in Anna University, Coimbatore and Hindustan University, Chennai. Currently he is guiding research scholars from the above mentioned universities. To know more about him, please log onto <http://www.travi.co.nr>.

Prof. R. Velayutham is currently Professor and Head with the Department of Computer Science & Engineering, Einstein College of Engineering, Anna University @ Tirunelveli. He graduated in Computer Science & Engineering from Madurai Kamaraj University. He obtained his masters degree from Madurai Kamaraj University. He is currently pursuing PhD from Anna University @ Tirunelveli. He has published 20 papers in International and National Conferences and 1 paper in National Journal. His areas of interest are Cryptography and Network Security. He has nearly 10 years of experience in teaching. He was previously with Noorul Islam University, Nagercoil.



9

**தமிழில் சிந்தனைத்திறன்
கணிணிச்செய்நிரல்கள்**



Design Of Intelligent Robotic Arm Writing Tamil For Physically Challenged

'A boon for physical ailment'

Seima Saki S, ECE

seima.skcet@gmail.com

Sabita Devi D, Mechatronics

sabi.skcet@gmail.com

Sri Krishna College Of Engineering And Tech., Coimbatore,

Manoj S, Computer Technology,

PSG College of Technology, Coimbatore s.manoj89@gmail.com

Mr.R.Vimalathithan, Senior Lecturer,

Sri Krishna College Of Engineering And Tech, Coimbatore athivimal@gmail.com

Abstract

Physically impaired people find hard to subsist in this challenging world. Here forth is advancement in technology to provide comfort to handicapped and visually disabled to convey their thoughts through pen and paper. Inventions made so far didn't prove a great amount of benefit to them. The working model proposed here is much novel and proves to be a well-worn technology. The central notion is to design a robotic arm that writes upon dictation. We use speech analysis techniques to convert it into text. Writing is brought by finding the mathematical equation of each alphabet, character or a number and converting into values that drives the arm. We use 3 degrees of freedom design for X,Y and Z axis motion. The values of the equations are stored in the memory of the micro controller and are sent to motors corresponding to the incoming text. Hence not only one but many styles of writing can be brought up by a single system by feeding their corresponding equation values. This not only assists disabled persons, but also supports people who need a proxy for their hand writing. This technique is much practical and shows best results.

Introduction

Researchers spend their valuable time in inventing novel methods to impede the disabilities of physically impaired. Previously developed devices were not that effective to make the needy feel relieved from all his necessities. Some tasks like signing the documents and writing in his own hand writing seemed to be impossible. Can a robot take up these jobs of a human...?

The technology invented here is to couple human voice with the mechanical arm. Speech is analyzed by software and converted to digital format. The hand writing is mathematically formulated as discussed below. Further, the system allows selecting upon different styles among the ones present in their memory. The rest of the paper is organized as follows: Section II presents a brief overview of

fundamental building blocks, Section III gives the overview of Hardware description, Section IV presents the construction details and Mathematical formula of text are discussed in Section V. Finally, Conclusion and Future work are presented in section VI.

Fundamental Building Blocks

Our aim is to make the arm trace the letters in accordance with the speech. Arm movement can be brought up with *pulleys and stepper motor*. For this, each alphabet, number or a special character is to be represented by a unique values, so called as '*fonts*', and fed into a memory. And motor carries out the activity in relation to those values. And we need a *microcontroller* to take up memory and control activities. We require a *speech analyzer* to convert human language into text and text is processed further into digital data. Microcontrollers work with digital inputs. So the speech has to be converted into digital string. To position and orient the robotic arm, manipulators of 6-DOF or 3-DOF can be employed. The DOF of the manipulator or the arm are distributed into sub assemblies of each axis.

Given below is the outline of technical blocks of the robotic humanoid hand.

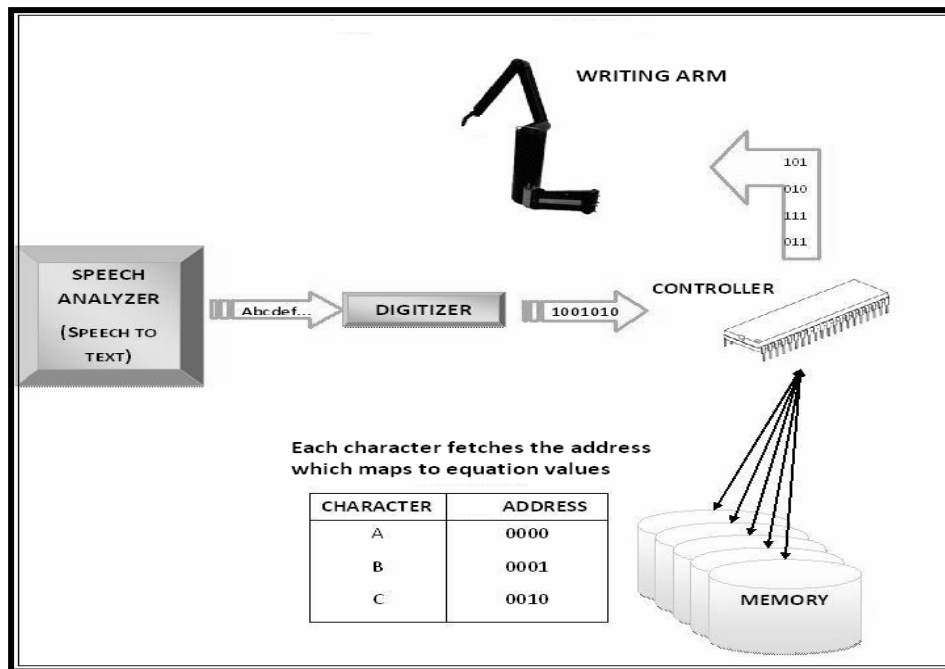


Figure 1: Basic Building Blocks-Overview

Hardware Descriptions

Micro Controller Unit:

We use the basic 8051 family micro controller Atmel 89C51. It has 4KB on-chip program space, 128bytes RAM, 2 timers, 4 I/O ports, 1 serial port and 6 interrupt sources. 8051 is a 8-bit processor meaning that the CPU can work on only 8 bits at a time. Data larger than 8 bits has to be broken into 8-bit pieces to be processed by the CPU.

Micro controller unit used here performs the task of matching the incoming text with already stored binary equivalents of the letters and alphabets. MCU makes decision of which data has to be written to the port where motors are interfaced. We require 3 motors for 3 Degrees of Freedom.

Stepper motor interface:

Stepper motor is widely used device that translates electrical pulses into mechanical movement. In applications such as hard disk, dot matrix printers, and robotics the stepper motor is used for position control. The most common stepper motors have four stator windings that are paired with a center-tapped. This type of stepper motor is commonly referred to as a four phase or unipolar stepper motor. The center tap allows a change of current direction in each of two coils when a winding is grounded, thereby resulting in a polarity change of the stator. The stepper motor shaft moves in a fixed repeatable increment, which allows one to move it to a precise position. This repeatable fixed movement is possible as a result of basic magnetic theory where poles of same polarity repel and opposite poles attract. The step angle determines number of rotations per minute and the delay given for each step determines the speed of rotation. Due to incremental steps, accuracy is high.

Pulley Mechanism for Navigation:

Movement of the arm can be brought about by using a manipulator that is fixed to a pulley belt. Pulley is operated by a stepper motor. The setup of stepper motor controlling the pulley with its shaft is shown in the figure. For x,y and z axis movement we use three pulleys with motors in it. Motion of one arm affects the other there by bringing up a relative translation over the entire 3 dimensional space.

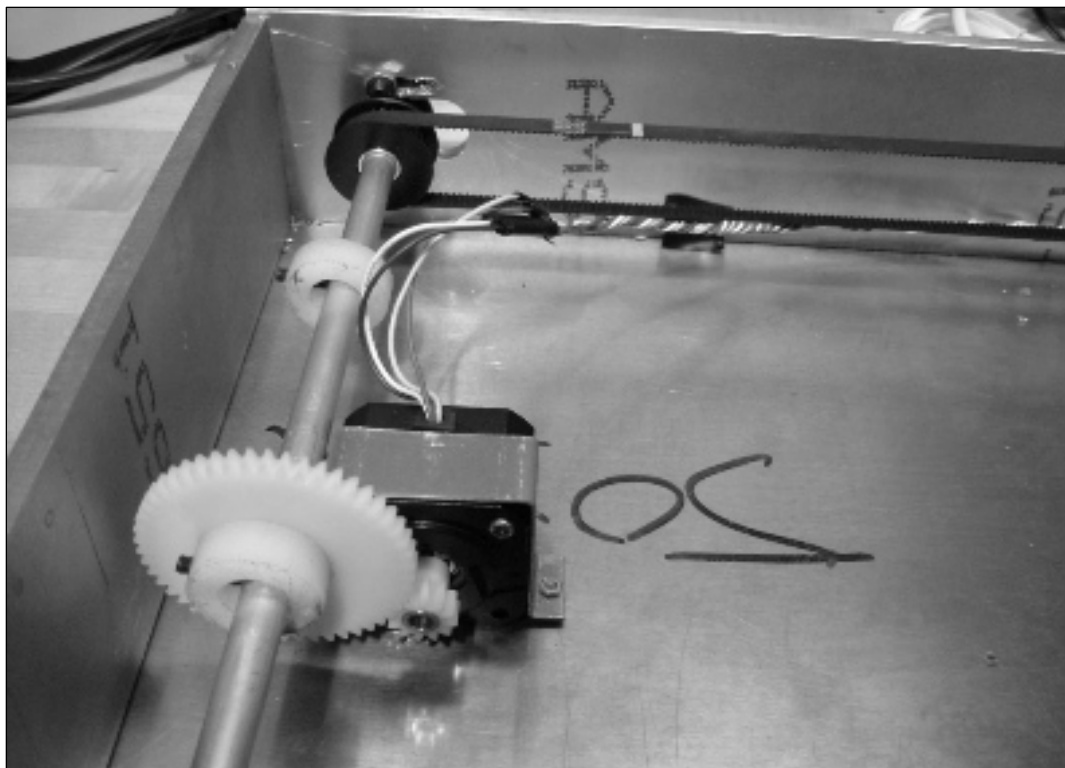


Figure 2: Stepper Controlled Pulley System

Construction

Module 1- Mechanical Perception:

3 DOF design:

To enhance robot hand dexterity, the robot should be designed to have a redundant number of degrees of freedom. In order to avoid ill poised nature and to minimize its inability to act as a human hand, we go for a 3 Degrees of freedom design. The over view of the design is shown in the figure. X,Y and Z are the Cartesian co-ordinates, q_1, q_2 and q_3 denotes vector of joint angles. L_1, L_2 and L_3 is the distance between two vector of joint angles. Movements in Z axis is used to lift the whole set up and place it in its origin for the start of a new page or a new line and movement in X and Y axis is for writing on the plane.

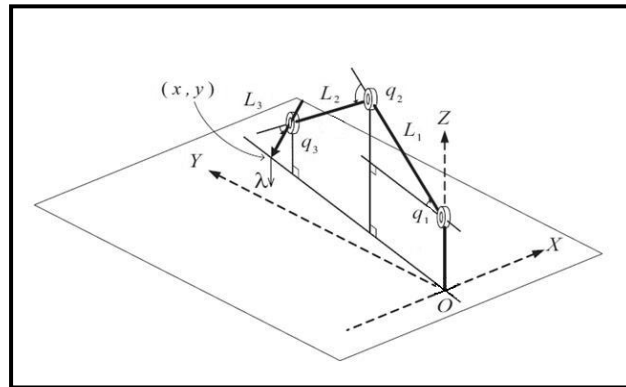


Figure 3: 3 DOF Design

Mechanical setup:

The entire built set up is shown in the figure. Pulley is supported by metal rods. As mentioned above we use three pulleys. Each of the pulley is driven by stepper motor, which gets signals from the micro controller for its movement. The strength of the pen on the paper is adjusted by varying the Z axis dimension. Also the speed of the stepper motor determines the accuracy of writing.

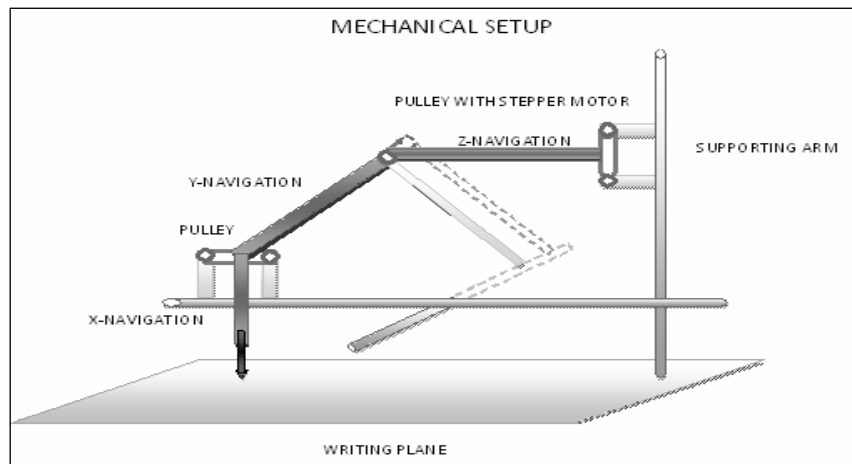


Figure 4: Mechanical Setup

Module 2- Programming Conception:

Speech to text:

The speech is converted to text using the third party software as mentioned above. The text data is converted to binary form and sent serially to MCU. Hence each word maps to certain set of binary values, i.e. set of letters in binary form. The pronunciations of words vary from person to person. Hence it is hard to differentiate between words having similar pronunciation. Bear can be interpreted as bare or vice versa. These types of verbal mistakes are much common with this type of systems and can be minimized by training the system continuously with most common words. If the software used is able to compare the words with the context of the sentence and then check for verbal errors, then it would be much efficient. As a key point, the text is converted to digital context and serially sent to the controller.

Text identification:

The values of each alphabet, character or number is already stored in heap of memory of the controller. This heap of values maps to set of locations which contains the values that are to be sent to the pen in order to bring about the letter. These values are nothing but the mathematical formulation of letters, i.e. the equations formed by those letters when endorsed in a 2 dimensional XY plane.

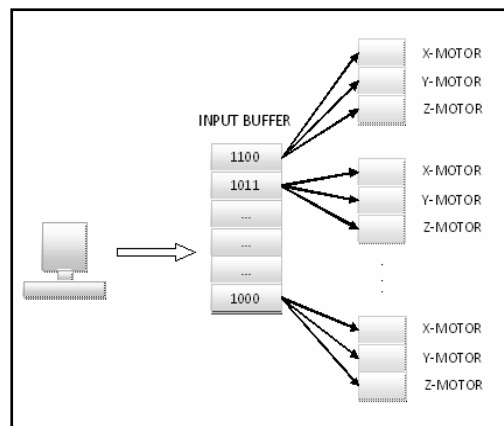


Figure 5: Text Mapping To The Memory

Mathematical formulation of text:

The entire text is imagined as combination of lines and curves in XY plane. Let the size of character be in the ratio of 1:2 (width : height) for English and 1:1 for Tamil. Hence each character is expressed in equations of x and y . For the word to be written, values of the entire axis have to be given simultaneously. For slanting lines, slope has to be specified and for curves its corresponding curvature values has to be specified. With this type of method, any character can be represented and be stored as specific fonts in the MCU memory. This allows multiple hand writings to be written by a single system.

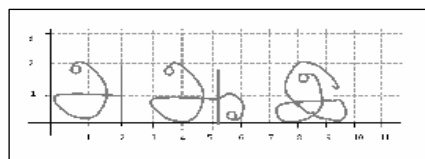


Figure 6: Graphical Outlook of a Sample Letter

Conclusion

The algorithm for formulation of the text was derived. Speech analysis was found tricky with many errors. But it was minimized when the system was trained with the common words. Mechanical arrangement was designed and the errors were significant while retrace. Yet it proved to be efficient. If this technology is released as a real-time product it would really be a boon for physical ailment.

References

- [1] R K Mittal, I J Nagrath: "Robotic and Control", Tata McGraw-Hill Publications.
- [2] Frederick: "Stastical methods for speech recognition" (chapter 1) -1997 MIT.
- [3] 8051 Microcontroller and Embedded Systems, Muhammed Ali Mazidi, Janice Gillispie Mazidi, McKinley
- [4] IEEE transaction on "Robottics and Automation", march 2003, vol. 10, Issue 1.
- [5] Third International conference on "Information technology and applications" 2005. Volume 2 issue 4 to 7 pages 21 to 24
- [6] IEEE transaction on "Speech And Audio Processing", July 2004 Volume 12, Issue 4
- [7] Ravi P. Ramachandran, Richard Mammone: "Modern methods of Speech processing", Kluwer academic publications 1995. pg no. 236.
- [8] Third international conference on "Information technology and applications", july, 2005. Vol. 2, issue 4.

LaaLaLaa - A Tamil Lyric Analysis and Generation Framework

Sowmiya Dharmalingam & Madhan Karky

sow_86@yahoo.co.in | madhankarky@gmail.com

Department of Computer Science & Engineering

College of Engineering Guindy, Anna University

Abstract

Over 1000 Tamil lyrics are being written every year. Tamil films, advertisements, private pop albums are the main contributors of Tamil lyrics. This paper presents LaaLaLaa, a framework for Tamil lyric analysis and generation. For the analysis part, a set of 2000 Tamil lyrics collected over a period of 50 years penned over different themes is mined for various statistics and morphological patterns. The analysed statistical data is then used for the generation of the lyrics. This paper explains the different components in the framework such as lyric generator, rhyme finder, lyric miner, template selector, morphological generator and WordNets for lyric generation. We introduce three scoring mechanisms for the properties of flow, rhyme and meaning of a lyric. These scoring methods are used to compare the quality of the generated lyric. Discussing the results, the paper concludes with open problems and future directions.

1. Introduction

Tamil is a vibrant language with a rich grammar, vocabulary, an inherent poetic flavour and music is synonymous with its culture. Tamil Lyrics have evolved in a dramatic fashion from the many historical poems of the Sangam literature including Ettuthogai and Pathuppaattu, narrative poem Silappatikaram, compositions of the Tamil Saiva saints such as Appar, Thirugnana Sambanthar, Manikkavaasagar and compositions of Tamil hymns known as Thiruppugazh to the medieval period philosophies dominated by songs to the present day's rap culture. There are about 1000 lyrics being written every year as private albums, jingles and as original soundtracks of mainstream movies.

In this paper, we propose LaaLaLaa, framework that can analyse lyrics for various statistics on words, their morphological patterns and generate lyrics in Tamil concordant to the input music and selected theme. Computational creativity is particularly very challenging, as it requires understanding and modelling knowledge, which almost cannot be formalized. Generating meaningful lyrics to a given tune and in a given domain can be treated as an optimisation problem that aims at maximizing the various features of lyrics such as meaning, rhyme and flow.

This paper is organised into four sections. The second section discusses a few contributions to lyric generation and how LaaLaLaa differs from the rest of the lyric generation systems. The third section presents the LaaLaLaa lyric analysis and generation framework and in detail discusses the various components of the system. The section also presents suggestions for three scoring mechanisms to

compare generated lyrics. The fourth section discusses results and concludes with open questions, ongoing research and future work.

2. Background

Several poetry generation systems have been developed in the past which broadly fall into the following categories - template based, generate and test approaches, evolutionary approaches and case-based reasoning approaches[6].

The Tra-la-Lyrics system[7] generates Portugese lyrics given a MIDI by calculating syllabic division and syllabic stress identification of words and the strength of each beat. It however doesn't handle the semantic aspect and produces completely random words and also doesn't use metric patterns that can serve as an exact template for the words.

The Automatic Generation of Tamil Lyrics for Melodies[12] identifies the required syllable pattern for the lyric and passes this to a sentence generation module which generates meaningful phrases that match the pattern. A corpus of poems and stories is used as a source of phrases. This system is constrained by the size of the phrases/patterns and generates phrases that are independent of the previous phrases which leads to lyrics that are meaningful in parts, but meaningless on the whole. Secondly the system presented in [12] generates rhyme based on maximum substring match and fails to make use of edhugai, moanai and iyaibu, the three rhyming patterns that are specific to Tamil language.

Other remarkable works in this domain include poetry generation in COLIBRI[3] and An expert system for the composition of formal Spanish poetry[5] which translate a user specified prose message into formal Spanish poetry, Hisar Manurung's McGonnagall[10] where the poem generation process is formulated as a state space search problem where a state in the search space is a possible text with all its underlying representation, and a move can occur at any level of representation, from semantics to phonetics. The POEVOLVE system which implements the architecture proposed by Levy[9] creates texts that satisfy the form specifications of limericks. The WASP[4] system splits a given block of text into shorter fragments to identify reference patterns and uses the words in the text to produce verses that match these patterns.

In his thesis, Manurung[10], defines a poem to be a text that meets three properties: meaningfulness, grammaticality and poeticness. Our system currently handles the grammaticality and meaningfulness aspects to a certain degree while work is going on to enhance these factors, to meet the poeticness factor (LaaLaLaa framework proposed in this paper uses edhugai monai and iyaibu properties to achieve this) and handle other limitations of existing systems.

3. LaaLaLaa Framework

The LaaLaLaa framework presented in figure 1, can be divided into four major subsystems. Music-To-Template, Lyric Analysis, Lyric Generation and Lyric Scoring are the four subsystems that constitute the LaaLaLaa framework. The following sections describe the subsystems in detail.

3.1 Music To Template

MIDI(Musical Instrument Digital Interface) is a format for representing music in digital devices. A MIDI file is the input to the system. The MIDI file is processed by a MIDI-To-ABC[1] converter. The ABC notes[8] are processed by ABC-To-Template converter which transforms the ABC notes to *Tamil Textual Place Holders*, we call them *Templates*. A template can be formed using combinations of a short vowel(tha | na), a long vowel(thaa | naa) and a consonant(n). An example line template is provided below.

It is also to be noted that a single template may match multiple patterns and a pattern is selected at random. This pattern tagged template forms the input to the Lyric Generation subsystem.

தானா <adjective+clitic> எனனை <noun+locative> எனனைஎனனை <verb+1st
 person + singular + future>

3.2 Lyric Analysis

Lyric Analysis is an offline system where a Tamil lyric corpus is analysed for various statistics and patterns. Lyrics collected from various sources are formatted by tagging them with appropriate headers and section tags(pallavi, anupallavi, saranam,..) The formatted lyrics are fed as input to a statistics miner and a pattern miner. A statistics miner analyses lyrics for statistics corresponding to length of sentences and words, length of co-occurring words.

n=Length(s)	AvgWPS(n)	AvgWLN(n)
5	2	3
10	3	3
15	4	4
20	5	4
25	6	4
30	7	4
35	8	4
40	8	6
45	8	6

Table 1 : Lyric Statistics

Table 1 provides some statistics obtained from our implementation of lyric miner over 2000 Tamil lyrics collected from multiple sources. The Length(s) denotes the character count for a given sentence. AvgWPS(n) denotes the average words per sentence for sentences with length n. AvgWLN(n) denotes the average character count of words for sentences with length n. It is to be noted that the average is computed over the lyric corpus. The template processor uses this information to efficiently split templates for word fitting.

wps	pattern	pattern
2	4	< Noun > < Pronoun + Sandhi + Clitic >
3	10	< Interrogative Noun + Clitic >< Pronoun >< Interrogative Noun + Clitic >
4	3	< Infinitive >< Verb + Past Tense + Verbal Participle>< Time >< Noun >

Table 2 : Lyric Pattern

A sample output from our pattern miner is provided in Table 2. wps denote the words per sentence and | pattern | denotes the number of times a particular pattern occurs. The pattern gives information on what

pos and morphological suffixes are used [2]. Pattern Selector module uses these results to tag templates for lyric generation.

3.3 Lyric Generation

The Lyric Generation Subsystem uses domain specific WordNets, Morphological Generator and a Rhyme Finder. The domain specific WordNet consists of words and their associations specific to a certain domain such as nature, love, history, geography, religion, zoology and more. Given a root word and appropriate tags, it can generate nouns and verbs. The information obtained from the Pattern Tagger is used to generate words according to the tagged pattern. The Rhyme Finder is used by the Lyric

தா <verb> னா <verb> னானா <noun+clitic>
 தா <verb> னா <verb> னானா <noun+clitic>

Generator to choose words that match one or more of the three Rhyme properties (edhukai, moanai, iyaibu). The edhukai property states that the second letter of two parallel words match. Two words are considered to be parallel if they occur in the same position of two different sentences. The moanai and iyaibu properties state the same for first letter and last letter to match respectively.

Let the above example be considered as two template patterns for parallel sentences sent to Lyric Generator.

The above given lyrics are two options generated by the lyric generator. The words poovae and theevae are not obtained from the WordNet as the WordNet comprises of root words only. The morphological generator uses the noun+clitic pattern for naanaa word template and chooses noun roots from WordNet

வா வா தீவே | vaa vaa theevae
 போ போ போவே | poa poa poovae
 வா வா தீவே | vaa vaa theevae
 போ போ பூவே | poa poa poovae

to generate morphological suffixes that meter-match the template.

Two words are considered to Meter match if the two words have same character length and every character of a word meter matches with the corresponding character of the other word. The meter matching of two words is again obtained from ancient Tamil grammar definition of maathirai aLav(u) (meter length). Rhyme finder minimises the search space for generating the second line by restricting the list of word to match edhukai monai or iyaibu properties. Choosing same length roots with same clitic or matching the first second or last characters of words enables choosing rhyming words.

3.4 Lyric Scoring

We propose a lyric scoring subsystem that takes an entire lyric as input and gives three different scores for the lyric namely flow score(f), rhyme score(r) and Meaning Score(m). Flow of a lyric can be modelled as a property influenced by the phonetic properties of words in the lyrics. Tamil alphabets are sensibly classified into short vowels, long vowels, soft consonants, hard consonants and mid consonants. We use this classification along with doublings in words to determine the score of a word, a sentence and thus a lyric. Rhyme score (r) analyses various rhyming patterns and rhyming properties mentioned in the

previous section determine a score for a lyric. Meaning score (m) is the challenge. We do not have any formal method to analyse if a given sentence is meaningful. If meaning can be a property that can be achieved through associations of words in a sentence and connection between sentences, then domain WordNets can be used to determine a meaning score for a lyric.

4. Conclusion and Future Work

The objective of this paper is to present the LaaLaLaa framework for lyric analysis and generation. Very few parts of the LaaLaLaa framework have been implemented. The word generator has been implemented in parts and a very small WordNet corresponding to nature is used for testing. A basic statistics and pattern miner modules have been implemented. Developing multi domain WordNets and implementing the meaning score will be a challenging task. Improving the speed of generation by building a special index based on meter length of words will be part of our future work.

Acknowledgements

We would like to thank Dr. T. V. Geetha and Dr. Ranjani Parthasarathi for providing valuable inputs to this paper. We also would like to acknowledge the contributions of Gayathri Lakshman, Prathyusha Senthil Kumar towards initial implementations of the LaaLaLaa framework.

References

1. Allwright, J., *The abcMIDI project* 2002, <http://abc.sourceforge.net/abcMIDI/>.
2. Anandan, R. Parthasarathi, and Geetha, *Morphological Analyser for Tamil*. ICON 2002, 2002.
3. Diaz-Agudo, B., P. Gervas, and P.A. Gonzalez-Calero. *Poetry generation in COLIBRI*. in *ECCBR 2002*. 2002.
4. Gervas., P. *Wasp: Evaluation of different strategies for the automatic generation of spanish verse*. in *Proceedings of the AISB00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science*. 2000. Birmingham, UK: Wiggins, G. (Ed.).
5. Gervas., P., *An expert system for the composition of formal Spanish poetry*. *Journal of Knowledge-Based Systems*, 2001. **14**.
6. Gervas., P. *Exploring quantitative evaluations of the creativity of automatic poets*. in *15th European Conference on Artificial Intelligence*. 2002.
7. Goncalo, H.R., et al. *TraLaLyrics: An approach to generate text based on rhythm*. in *Fourth International Joint Workshop on Computational Creativity, IJWCC'07*. 2007. London.
8. Gonzato., G., *The abc plus project*. 2003, <http://abcplus.sourceforge.net>.
9. Levy, R.P. *A computational model of poetic creativity with neural network as measure of adaptive fitness*. in *ICCB-01 Workshop on Creative Systems*. 2001.
10. Manurung, H.M., *An evolutionary approach to poetry generation*. 2004, University of Edinburg.
11. Manurung, H.M., G. Ritchie, and H. Thompson. *A flexible integrated architecture for generating poetic texts*. in *Fourth Symposium on Natural Language Processing*. 2000. Chiang Mai.
12. Ramakrishnan, A., S. Kuppan, and S.L. Devi. *Automatic Generation of Tamil Lyrics for Melodies*. in *NAACL HLT Workshop on Computational Approaches to Linguistic Creativity*. 2009. Colorado.

Language Independent Emotion Recognition System for Web Articles Using NLP Techniques

D. Mahendran,

PG Scholar | dmahendranme@gmail.com

S. Gunasundari,

Senior Lecturer | gunaannauniv@gmail.com

Prof. B. Rajalakshmi,

HOD | velammalcseraji@yahoo.co.in

Department of CSE, Velammal Engineering College, Chennai, India.

Abstract

In recent years, suicide of college students has been a universal phenomenon in the world. And the phenomenon has become more and more severe because of the complex and drastic competitions. With the popularization of internet and the development of information processing technologies, a lot of people have established their own blog websites to write down their experiences and express their feelings at times. It will be very helpful if the computer can recognize the emotions expressed in blog pages automatically. And then it will be convenient for teachers or psychological consultants to monitor the affective information of college students and take measures for the depression prevention when necessary.

Owing to the advances in Affective Computing and Natural Language Processing, researches have begun to pay more attention to the emotion recognition in NLP all over the world. This leads to construct an Emotion Recognition System. It is based on the lexical contents of words and structural characteristics of blog articles. The emotion recognition system is composed of five modules: data collection, data processing, morphological analyzing, emotion tagging, and emotion computing. This algorithm independently works for web articles in any language. We checked articles in English and Tamil and the response of the system is good.

Keywords: Suicide, Depression, Natural language processing, Emotion classification, Emotion recognition

Introduction

In recent years, suicide of college students has been a universal phenomenon all over the world. Furthermore, the number of college students who suicide themselves has been increasing with amazing speed in a worldwide scope, around 100 million persons every year. In the year of 2003, World Health Organization (WHO) appointed September 10th as "World Suicide Prevention Day (WSPD)" to give a caution to the society. And it has adopted many suicide prevention programs and projects to avoid unnecessary death. At the same time, the problem of mental health among college students has raised great attention from more and more psychologists, educators and people in all the other professions.

In order to intervene in the crisis, psychological consultation centers have been set up almost in every university or college. However, some college students are afraid to talk about their private experiences with other people face to face, and refuse to consult psychological consultants when they meet troubles.

In late decades, the popularization of internet and the development of information processing technologies have greatly changed the communication ways of mankind. More and more people have established their own blog websites over the internet where they write down their experiences, put forward their opinions, and express their feelings at times. While some people feel embarrassed to confide their troubles to others face to face, they may feel free to express their emotions in blogs without any pressures.

It can be anticipated that in the near future, almost everyone in the world especially every college student will possess his or her own blog website. At that time, it will be very helpful if the computer can recognize the emotions expressed in blog articles automatically.

In this case, an Emotion Recognition System is needed. If the system detects that some college student has been in a blue mood for continuous days (for example, for continuous three days), it is recommended that the teacher should pay more attention lately and have a talk with him or her regularly so that the depression can be treated in time. That's why this research has been taken on.

In this paper, we outline a new approach to recognize the author's emotion from his or her blog articles. Based on the approach, an Emotion Recognition System has been constructed. The results of our experiments on blogs prove the feasibility of the means.

System Structure

The emotion recognition system is composed of six modules: data collection, data preprocessing, dictionary creation, morphological analysis, emotion tagging, and emotion computing shown in figure 1.

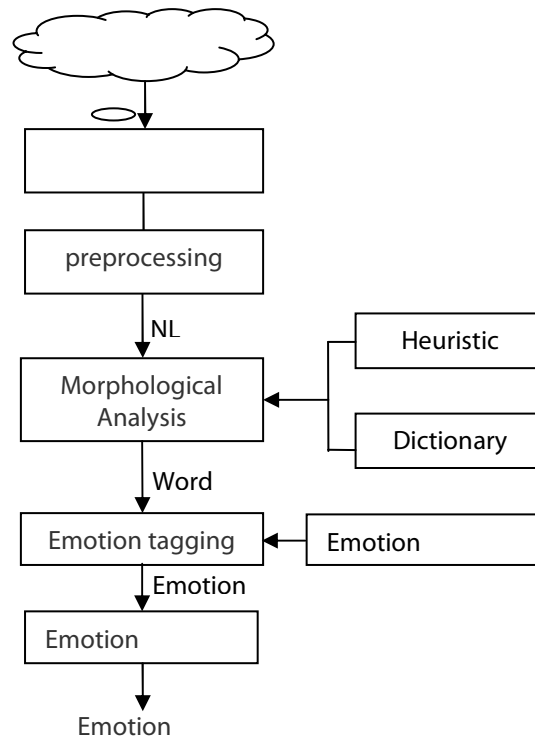
In the following we will introduce the flow of how the system works step by step. Firstly, the blog articles are collected from various authors with their emotion category mentioned. The collected articles are preprocessed to remove unwanted data from the blog articles. Then the preprocessed articles are subjected to morphological analysis. Here in morphological analysis the articles are split into sentences then the sentences are split into words for comparing with the emotion dictionary to find the emotion category. In emotion tagging, emotion category for each words and sentences are assigned.

Finally, with the emotion category of each sentence, the Emotion Computing module will compute the weight value of emotion categories of all sentences in a blog article according to the blog structure rules, and output the emotion result for the whole article.

Data Collection

Our emotion recognition system will work effectively on group of sentences. That's the reason we go for blog articles where people express their emotions in blogs without any pressures. In our country the blogs are not that much famous. This is the issue we faced on collecting the data. The personal blogs are popular in US, China, Japan which is not that much popular in our country. What we did is, we created a proxy to access the US websites for English blogs. But in Tamil, we got only few web articles.

Figure 1: Framework for the emotion recognition system



Around **300 English blog articles** were collected from the internet among various authors.

Data Preprocessing

The collected blog articles contain images, videos etc. Raw data cannot be used directly in our emotion recognition system. So that, the data has to be preprocessed. Here in data preprocessing the unwanted elements are removed from the articles like HTML tags, images, videos and other multimedia files are removed. HTML Parser is used for data preprocessing. The HTML Parser module analyzes the blog webpage that has been downloaded, and extracts each blog article in the webpage. Then it divides the blog article into paragraphs and sentences which are stored for further processing. Delimiters used for separating the sentences are .(dot), ?(question mark) and !(exclamatory mark). Other special characters like -, @,#,\$ etc are removed from the articles. And for Tamil articles, the format is changed to Unicode for easier implementation.

Morphological Analysis

Morphology is the identification, analysis and description of the structure of words (words as units in the lexicon are the subject matter of lexicology). While words are generally accepted as being the smallest units of syntax, it is clear that in most languages, words can be related to other words by rules. Here in our system we have created the own morphological analyzer. It contains Heuristic rules, English and Tamil dictionaries (noun dictionary, verb dictionary, adverb dictionary). User interface of one of the dictionary is shown in figure 2. Binary search tree algorithm is a node based binary tree data structure which we have used in these dictionaries. Therefore the searching efficiency will be more. The heuristic

rules are rules to extract the root words from the sentence. For example: The word ‘happily’ is an adverb. Root word for ‘happily’ is ‘happy’. Likewise ‘anju’ is a root word for Tamil word ‘anjinen’

For each natural language sentence, the Morphological analyzing module will analyze the lexical characteristics of the sentence and extract words from it based on the rule database according to analyzer rules. The words are used for further processing. Morphological analysis part has been made of automatic.

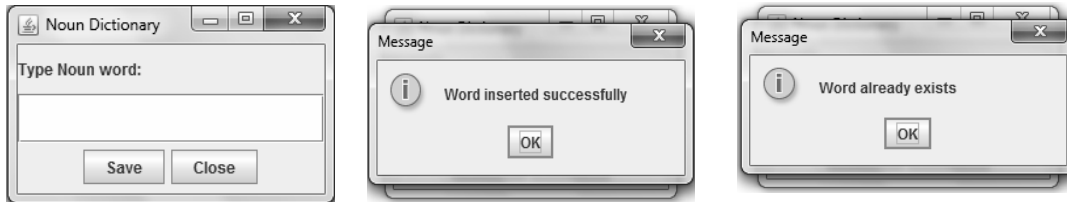


Figure 2: User interface of the Noun dictionary

Emotion Dictionary

Emotion dictionary is constructed which is to be used in emotion tagging module. In emotion dictionary, the data structure, Binary search tree algorithm is used to improve the searching efficiency. Emotion dictionary is made as semi-automatic. User interface of the emotion dictionary is shown in figure 3. Here two separate emotion dictionaries are used for English and Tamil languages. Numerous words have been inserted in the emotion dictionary with their appropriate category. It contains three fields, word, category and weightage. The weightage field stores the weight values for each word. Weight values of each word play a vital role in emotion recognition system. Emotion category is found out based on weight values.

In the past, emotion has been simply divided into two categories “pleasure” and “displeasure” which are too ambiguous to assess rich emotions of human. Ekman has defined 7 universal affective categories [2] based on unique facial expressions which seem still less while applied into a practical system. In contemporary Chinese, 39 emotional categories are specified for vocabulary [3], nevertheless part of which is seldom used in daily communications.

Table 1: Emotion vocabularies for Tamil

Emotion category	Happy	Angry	Sad
vocabulary	Santhosham	Sinam	Thukkam
	Subam	Veruppu	Thuyaram
	Sugam	Padhatram	Varutham

In our Emotion Recognition System, we classified emotion based on the affective categories defined by Ekman plus some important categories found in blogs with high frequency of use. They are totally 26 categories: happy, sad, fearful, disgusted, angry, surprised, love, expectant, nervous, regretful, praiseful, shy, respectful, proud, impatient, doubtful, hateful, grievance, critical, depressed, exited, thankful, annoyed, scornful, haughty, envious. When there is no emotion expressed in a blog, we named the mental state as “neutral”. Some examples of the emotional vocabularies for Tamil are listed in Table 1.

Emotion Tagging

In morphological analysis the article is split into sentences and words. Those extracted words are passed to emotion tagging module. The words are compared with the emotion dictionary. If the corresponding word is present in the emotion dictionary then the weight values and emotion category is assigned to that word.

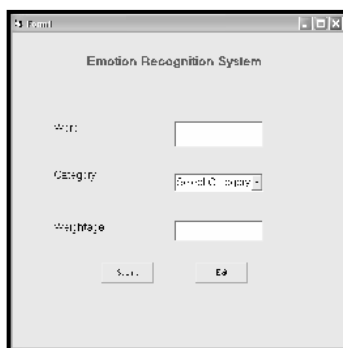


Figure 3: User interface of the emotion dictionary

Emotion Computing

Finally, with the emotion category of each sentence, the Emotion Computing module will compute the weight value of emotion categories of all sentences in a blog article according to the blog structure rules, and output the emotion result for the whole article.

Let us suppose that there are j ($j=1,2,3,\dots,m$) sentences in an article. For each sentence j , there is a emotion weight value W_j assigned. E_a is one of the 26 emotion categories ($a=1,2,\dots,26$). E_a with maximum sum of all the corresponding W_j is the emotion of the whole article.

Experimental Results

Emotion recognition system is implemented in Java language. And we have attained the expected results. It is still an issue for the evaluation. By now there is not a good evaluation method generally accepted. Since different people may have different opinions even on the same text, it is common for them to give different evaluations of emotion manually. In our experiments, we carried close test based on the manual evaluation of emotions as a standard set which is judged by evaluators in advance.

Close Test: The close test is carried out on several blog articles. Our algorithm of emotion computing is realized in our system. And the evaluation is to count the number of correct prediction by the system compared with the standard set. We checked articles in English and Tamil and the response of the system is good.

Conclusion and Future Work

Recently, the number of college students who suicide themselves has been increasing with amazing speed in the world. The problem of mental health among college students has raised greater attention from a lot of people in many professions. With the popularization of internet and the development of information processing technologies, more and more people have established their own blog websites. It will be convenient for teachers or psychological consultants to monitor the affective information expressed in blog articles in order to prevent the depression of students. The advances in Affective

Computing and Natural Language Processing enable the emotion recognition from blog articles automatically.

In this paper, we have outlined the approach to develop an Emotion Recognition System. Firstly we decided the classification of emotions, and introduced the model of the System. Based on the lexical contents of words and structural characteristics of web articles, a method have been proposed for emotion computing. The experiments for testing have been carried out. The approach was proved to be feasible and pointed out the future's direction of the research. We find that some parts need to be improved in the future: The first task to improve the emotion dictionary by expanding the vocabularies and arranging them better, for it plays a very important role in the emotion sensing. Since different people have different styles in writing blog articles, in the future we will choose more web articles of different authors for the algorithm analysis and system test, which are supposed to achieve a better performance. Currently we are developing a system to find emotions from Hindi language blogs. We are in research of developing a language independent emotion recognition system for all languages. That is, identifying an emotion from all language blog articles by the help of Unicode.

References

- [1] P. W. Picard, E. Vyzas, J. Healey, "Toward Machine Emotional Intelligence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No.10, pp. 1175-1191, October 2001.
- [2] P. Ekman, W. V. Friesen, P. Elsworth, "Emotion in the Human Face", Cambridge University Press, London, 1982.
- [3] X. Y. Xu, J. H. Tao, "Emotion Dividing in Chinese Emotion System", the 1st Chinese Conference on Affective Computing and Intelligent Interaction (ACII'03), pp. 199-205, Beijing, China, December, 2003.
- [4] K. Matsumoto, J. Minato, F. J. Ren, S. Kuroiwa , "Estimating Human Emotions Using Wording and Sentence Patterns", Proceedings of the 2005 IEEE International Conference on Information Acquisition, pp. 421-426, 2005.
- [5] D. Kulic, E. A Croft, "Affective state estimation for human-robot interaction", IEEE Transactions on Robotics 23 (5), pp. 991-1000, Oct. 2007.
- [6] Xiaoxi Huang, Yun Yang, Changle Zhou, "Emotional Metaphors for Emotion Recognition in Chinese Text", Springer Berlin / Heidelberg publication on Affective Speech Processing, pp. 319-325, November 15, 2005.
- [7] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J.G. Taylor, "Emotion recognition in human-computer interaction", IEEE Transactions on Signal Processing Magazine, Vol. 18, No.1, pp. 32-80, January 2001.
- [8] Y. Zhang, Z. M. Li, F. J. Ren, S. Kuroiwa, "Semi-automatic Emotion Recognition from Textual Input Based on the Constructed Emotion Thesaurus", Proceeding of NLP-KE'05, pp. 571-576, November 2005.
- [9] Ye Wu, F. Ren, "Emotion recognition based on negative words and pattern matching for Chinese negative sentences", Proceeding of NLP-KE'08, pp. 1-5, October 2008.
- [10] H. Li, N. Pang, S. Guo, H. Wang, "Research on textual emotion recognition incorporating personality factor", Proceeding of ROPIO'07, pp. 2222-2227, December 2007.
- [11] <http://www.tamilcafe.net/>

An Integrated Intelligent Framework for automatic story generation in Tamil

G.V. Uma

Assistant professor

Dept of computer science Engineering

Anna University Chennai, Chennai.

gvuma@annauniv.edu

Abstract

Story is a description of a chain of events told or written in prose or verse. It is an interesting way to transfer knowledge from one person to other in the form of narrated sequence of events. The sequence of events is arranged in a chronological order to convey the message to other. An intelligent framework is developed using ontology to benefit the system in order to generate the story automatically and to reason the system generated-stories based on their conceptual consistency and validity. Ontology has the significant role in constructing semantic stories by the system. Ontology is a formal explicit shared conceptualization, which contains domain knowledge for the story construction. This research work developed a framework for automatic story generation in Tamil.

Introduction

Everyone in the world have their own interest to read the stories and like them very much. Stories are naturally has their own way of attraction from children to old age people. Children learn their moral and social obligations in the form of stories narrated to them by their guardians and peers [1]. The basic characteristics of human beings can be explained through stories to the youngsters to inspire them. Hence, they play a vital role in everyone's life and their traits can be shaped well, based on the characters in the story. For example, Mahatma Gandhiji was influenced by the story "Harichandra" for speaking the truth in any kind of difficulties, and the "shravana story" for obedience to his parents. People can gather a lot of information from the story based on their perceptions.

The ASG system tries to make the computer and Artificial author' to construct new stories dynamically. The story is a natural verbal description of objects / human beings, their attributes, relationships, beliefs, behaviors, motivations and interaction. It is a message that tells the particulars of an act, and an occurrence or course of events are presented in the form of writing. It can also be described as follows:

- It is a description of the sequence of events
- It is a piece of fiction that narrates a chain of related events
- It focuses on only one incident, has a single plot, a single setting, a limited small number of characters, and covers a short period of time.

The automatic story generation system helps to generate a variety of new stories dynamically as per the reader's interest. The user has the choice to select any characters, locations, settings for generating a story. The theme may be conceived either automatically or by the user; similarly, existing stories can be revised to remake new ones. Any kind of story can be restructured and reasoned based on the author's

requirement and specification. Ontology provides full support for story generation by preserving the concepts, attributes and the relations among them, which help for semantic reasoning. The language generator helps to generate suitable sentences with beautiful words for the construction of new stories.

This story generation framework enables the user, to act as reader as well as the author of the story. They can generate stories based on their wish and with necessary ingredients like characters, settings, location, food, etc. A simple story is started with initial situation and which has an active element to precede the story interestingly and finally comes to final situation. The generated stories are helpful for kids to know about the world knowledge.

Story generation framework comprises of three levels of computation such as theme conception, story generation and semantic reasoning. Ontology plays a major role in each of these three phases. Ontology is a formal explicit specification of shared conceptualization [2]. It can be expressed as the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary for a specific domain. The ontology is constructed based on the components of story domain. It involves characters, location settings like forest, palace, home and etc. The ontology possesses the events and the order of events/ functions /activities leads the story generation system. Even though constructed ontology is domain specific, it can be updated, modified, reengineered, reused for other purposes also. The main reasons [3] for using ontology for story generation system is

- To share common understanding of the structure of information among people or Software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

The structure of the paper organized as follows: Section 2 discusses the related work and Section 3 provides the framework for automatic story generation. Section 4 devotes in the role of ontology in automatic story generation. Section 5 focuses on the role of ontology in story generation and reasoning. Section 6 discusses the conclusion and future works.

Related works

There are different types of story generators available for the purpose of automatic story generation and their evolution is described below.

Propp [4] discussed the story generation as; a tale is a whole that may be composed of thirty-one moves. A move is a type of development proceeding from villainy or a lack, through intermediary functions to marriage, or to other functions employed as a denouement (ending). One tale may be composed of several moves that are related between them. One move may directly follow another, but they may also interweave; a development, which has begun pauses and a new move, is inserted. Bailey [5] described an approach towards automatic story generation based on the twin assumptions that it is possible for the generation of a story to be driven by modeling of the responses to the story of an imagined target reader, and that doing so allows the essence of what makes a story work (its 'storiness') to be encapsulated in a simple and general way.

Charles, F et al [6], presented results from a first version of a fully implemented storytelling prototype, which illustrates the generation of variants of a generic storyline. These variants result from the interaction of autonomous characters with one another, with environment resources or from user intervention.

Dimitrios N. Konstantinou et al [7] discussed the story generation model HOMER. It receives natural language input in the form of a sentence or an icon corresponding to a scene from a story and it generates a text-only narrative apart from a story line and it includes a plot, characters, settings, the user's stylistic preferences and also their point-of-view.

Riedl et al [8] had provided planning algorithm for story generation. The story planners are limited by the fact that they can only operate on the story world provided, which impacts the ability of the planner to find a solution story plan and the quality and structure of the story plan if one is found, but which lacks semantics.

George miller [9] provided a wonderful environment to have the collection of words and their synonyms and they are put together to form a lexical database. It helps to retrieve the meaning for any kind of word in any language.

This framework helps to generate stories automatically and to overcome the semantic lacking of stories which can be resolved by reasoning the stories in a systematic and efficient way using the efficiencies of ontology.

Framework Automatic Story Generation

This framework for automatic story generation shown in Figure 1 and it helps to provide a systematic approach to generate and analyze stories both syntactically and systematically. This framework mainly divided into five phases such as Theme conception, Sentence generator, Parser, Syntactic checker and Semantic checker.

Theme Conception -

There are two main approaches for conceiving theme such as:

- **Static conception** - is a predefined order of events to describe the flow of a story.
- **Dynamic conception** - refers to the order of events which are selected and organized during the phase of story generation

Set of built-in themes are available in the theme repository; the user can select the necessary theme for the story based on the characters involved. Also, a new theme can be conceived using events from the repository.

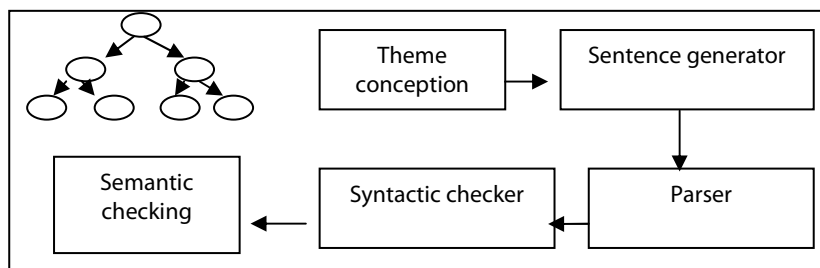


Figure 1 – Framework for automatic story generation

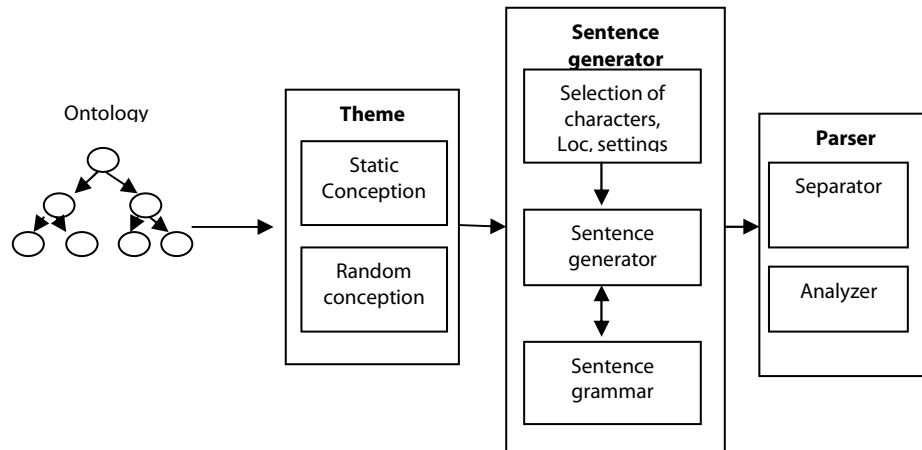


Figure 2 – Detailed Framework for Sentence Generator

Figure -2 depicts the detailed framework for generating simple sentences .The inputs of the Sentence generator are sentence type, sentence grammar and necessary terminal values. Using the sentence type, the corresponding sentence grammars and their production rules are retrieved from ontology. The necessary terminal values are checked with given production rules. There is no standard sentence structure for Tamil. The following grammar rules were framed and based on the rule sentence structures is obtained [11] [12].

1. NC --- > adj N / N / ADJC N / NNC
2. VC --- > adv V/ adv rp1 / ADVC V / vp1 / V
3. NNC --- > S con
4. ADJC --- > NC VC
5. ADVC --- > (NC)* vp1
6. S --- > (NC)* (VC)

Figure 3 – sample grammar for sentence generation

Figure 3 depicts the sample grammar for sentence generation in Tamil which is used for stories. By applying if then – rules, suffixes are updated with root words. Parser contains the two phases namely separator and analyzer. Separator helps to separate the story into story segments called sentences. The separation of sentences helps to check the sentences by its form of sentence structure and the meaning of the story. Analyzer used to identify the noun, verb, settings, location and etc from the story which helps in semantic validation.

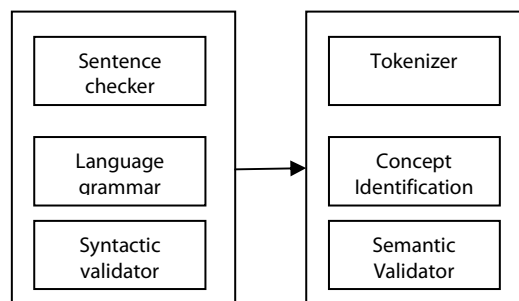


Figure 4 – detailed Framework for Semantic Reasoner

Figure 4 depicts the detailed description about the semantic reasoner which divides into two such as syntactic validation framework and semantic validation framework. Syntactic validation framework helps to check the sentence structure of the sentences based on the sentence grammar whereas semantic checker detects the meaning of sentences whether it is valid or invalid..

Role of ontology in ASG

Nowadays ontologies have significant role in information processing. They are important for story generation, which holds the various concepts that are relevant to the story domain. Initially, the ontology is built with minimal knowledge, in later stages; ontology can be extended whenever new concepts are introduced. One of the primary purposes of constructing ontology is to provide a standard, unambiguous representation of a particular domain of knowledge.

OWL has the language expressive representative formalism and reasoning power. Because OWL was derived from DAML+OIL, it can take advantage of the existing reasoning algorithms in Description Logics (DL). The semantic of OWL allows us to define a ranking function that distinguishes multiple degrees of matching. There are three types of matches: exact match where the concept to be found is found, plug-in match where the concept to be found is more specific than the concept in ontology, and subsume match where the concept to be found is more general than the concept in ontology. The scoring function of matching degree is given below [10]:

Exact Match > Plug in Match > Subsume Match

This matching degree helps to identify the semanticness of the generated story. The basic properties of the Lion is

Living being → animals → wild → Lion (king, legs, anger, roar, kill)
→ Domestic → Rat (small, legs, frightens)
→ Bird → crow (black, legs, fly)

If the generated sentence states that ' சிங்கம் பறந்தது. '(Lion flew) means the semantic checker detects that 'fly' action cannot be performed by animal. So the basic properties of Lion are preserved. It helps to improve the quality of the story. Consider another sentence, " எலி சிங்கத்தை கொன்றது. " in this sentence, the domestic animal 'mouse' killed the wild animal 'Lion'. In reality, it is not possible and then the system detects it, based on the strength of the animals which are categorized in ontology. The semantic reasoner detects and corrects the sentence as, So, the எலியை சிங்கம் கொன்றது. basic properties of Lion and rat are preserved. The sample story generated by the system is given in the Figure 5.

சிங்கம் காட்டில் தூங்கியது.
சிங்கத்தை எலி எழுப்பியது.
சிங்கம் கோபம் கொண்டது.
எலியை சிங்கம் கொன்றது.

Figure 5 – Sample story generated by the system.

Results

The above framework has been implemented in Java and uses the Java JDBC to connect to MySQL server and retrieve the relevant values. GUI is developed using Java Swing. The necessary basic concepts and their attributes are retrieved for semantic checking. The framework has been built and tested for various stories. The system parses the sentence to the framework for the reasoning purpose. Charles [6] proposed a set of factors that are considered to check the quality of the story. These factors are utilized to check the quality of stories by reasoning. Based on the above factors, the generated stories are given to a group of people to give their opinions about the generated stories with following scaling factors and with above said features. Excellent - 5; V.Good - 4; Good -3; Fair - 2; needs improvement - 1;

Table1: Factors for Assessment of Story

S.no	Factor	Describes
1	Overall	How is the story as an archetypal fairy tale?
2	Style	Did the author use an appropriate writing style?
3	Grammaticality	How would you rate the syntactic quality?
4	Flow	Did the sentences flow from one to the next?
5	Diction	How appropriate were the author's word choices?
6	Readability	How hard was it to read the prose?
7	Logicality	Did the story seem out of order?
8	Believability	Did the story's characters behave as you would expect?

Table2: Results of Story Assessment

S.no	Parameters	Before reasoning (max = 5)	After reasoning (max = 5)	Improvement
1.	Overall	4.3	4.7	1.09
2.	Style	4.0	4.5	1.13
3.	Grammaticality	4.1	4.8	1.17
4.	Flow	4.2	4.5	1.07
5.	Diction	2.8	3.4	1.21
6.	Readability	3.7	4.0	1.08
7.	Logicality	3.6	4.1	1.14
8.	Believability	4.4	4.7	1.07
	Average	3.885	4.33	1.12

The Table-1 shows the assessment value of story. The Table - 2 depicts the calculated values for the stories before reasoning and after reasoning. After reasoning, the quality of the story is improved on an average of 22.4 percentage better than the before reasoning. The believability factor and overall content have very good feedback among other factors. The other factors like style of the story, grammar content in the story, flow of the story are the good factors in the next level and also the other good factors in the next level are diction, readability, and legibility. Figure 6 depicts results of story assessment of stories.

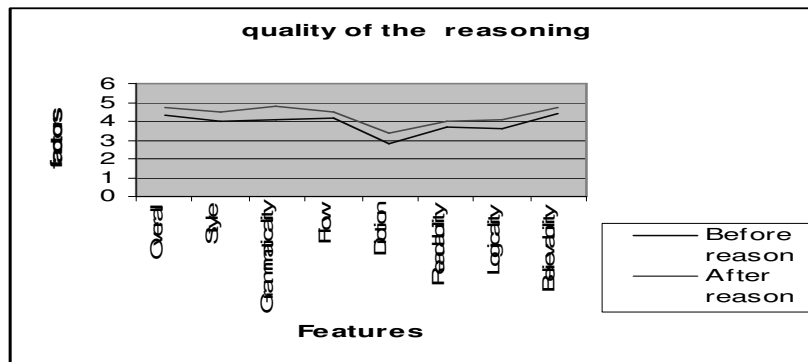


Figure 6 - Results of Story Assessment

Conclusion

Thus the framework helps to generate a simple short story in Tamil with simple sentences and semantic reasoning. It also proves that the framework is efficient, by comparing the results of the generated stories with before and after reasoning. In future, the project work can be extended to generate medium size stories, novels and picture based stories. Also, ontology can be extended with more number of concepts, attributes and their relation. Similarly, Reasoning can be extended to complex type sentences too.

References

- [1] Bilasco, I.M., Gensel, J., Villanova-Oliver, M.: STAMP: A Model for Generating Adaptable Multimedia Presentations. *Int. J. Multimedia Tools and Applications*, Vol 25 (3) (2005) 361-375.
- [2] Thomas R. Gruber Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli, (Eds.), *International Workshop on Formal Ontology*, Padova, Italy. Revised August 1993. Published in *International Journal of Human-Computer Studies*, Volume 43 , Issue 5-6 Nov./Dec. 1995, Pages: 907-928, special issue on the role of formal ontology in the information technology
- [3] Natalya F. Noy and Deborah L. McGuinness. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001
- [4] Propp, V "Morphology of the Folktale", University of Texas Press, 1968.
- [5] Paul Bailey, "Searching for storiness: Story generation from a Reader's perspective" Symposium on Narrative Intelligence, AAAI Press, 1999.
- [6] Charles, F.; Mead, S.J.; Cavazza, M. "Character- driven story generation in interactive storytelling" *Virtual Systems and Multimedia*. Proceedings. Seventh International Conference on Virtual Systems and Multimedia. 25-27 pp no: 609 - 615, Oct. 2001
- [7] Dimitrios N. Konstantinou , Paul Mc Kevitt , " HOMER: An Intelligent Multi-modal Story Generation System" Research plan. Faculty of Informatics, University of Ulster, Magee, Londonderry, 2002.

- [8] Riedl, M. and Young, RM, "Open-World Planning for Story Generation" Proceedings of the 19th International Joint Conference on Artificial Intelligence. California USA 2004.
- [9] George Miller , "WordNet. An electronic lexical database". Edited by Christiane Fellbaum, with a preface by . Cambridge, MA: MIT Press; 1998. pp 422
- [10] Ong Siew Kin, Tang Enya Kong, "Conceptual Modeling and Reasoning using Ontology "National Computer Science Postgraduate Colloquium 2005 (NaCSPC'05).
- [11] Saravanan, K., Ranjani parthasarathi, Geetha.T.V., "*Syantactic Parser for tamil*", Tamil internet, 2003.
- [12] T. Mala , T.V. Geetha, 'An Intelligent System for Picture based Tamil Sentences' International forum for information technology in Tamil and Insitute of Indology and Tamil studies, Germany, 2009.

Evaluation of Tamil Descriptive Passages using Concept Maps

Mahalakshmi G.S

mahalakshmi@cs.annauniv.edu

Sendhilkumar S

thamaraikumar@cs.annauniv.edu

Shankar B

shanloyola@yahoo.co.in

Department of Computer Science and Engineering,
Anna University, Chennai 600025, India

Abstract

E-learning has changed the way of education. With the wide spread applications of E-Learning technologies to education at all levels, we need to focus on learner-centric knowledge management in order to complement the conventional learning system. This paper examines the pioneering application of concept mapping as a follow-up study strategy for learning from text. With the great advancement in language processing tools in Tamil, we have proposed a novel mechanism which will evaluate Tamil descriptive answer passages based on the pre-stated prose. The idea is achieved by generating concept maps equally from prose and descriptive answer passages, and thereby evaluating the contents via generation of concept maps.

1. Introduction

With the wide spread applications of information technology to education at all levels, we need to focus on learner-centric knowledge management in order to complement the conventional learning system. To enhance self-learning and assessment especially in learning subjects that require high knowledge retention, application of concept maps to promote education has been a proven success worldwide.

1.1 Academic Significance

The first difficulty someone who attempts to comprehend a text faces is to understand what it is all about. That is, to grasp the global sense of the communication, understand its elements and the relationships among them. In this context, the student may understand some of the concepts involved in the definition. These concepts are linked by words forming whole sentences that seem to make sense. However, trying to understand the overall conceptual structure is more difficult. It is probably easier for many students to grasp a whole sense of the concept frame of reference when faced with a graph like the structure. This is due to the powerful visual effect that a graph has in order to facilitate understanding of a concept or a conceptual structure.

1.2 Contribution to Knowledge

Concept Maps **help to improve understanding of a given subject** and facilitate building student's own knowledge, as long as the student has the opportunity to use, criticize, analyze, question or improve expert's maps or Concept Maps generated by his own peers.

The implementation of concept maps in the classroom allows both the teacher and the student **discovering and describing meaningful relations among the concepts** object matter of the study, making it possible to **create connections** between them and the context in which activities are developed.

The concept map helps the learners to have a **better overview of the course** and what aspect he/she should pay attention. Concept maps constructed are very useful for teachers as an **evidence** of the way as each one of the parties involved in the process assumes his/her own learning. From their **follow-up and analysis**, experiences can be designed to help their students overcome weaknesses or to reinforce strengths acquired in learning process.

This motivated us to apply technological advancement and research in contributing better methodologies for education with a special focus on students whose medium of instruction is any regional language (for our study: Tamil) other than English. Although concept maps have been proven to be a successful resource for grade improvement in abroad, their use is little explored in Indian education. The detailed literature analysis conducted for the same revealed the fact that almost no work reported the development and use of concept maps to promote education in regional language – Tamil. In this context this paper concentrates on development of concept maps that eliminates the need for memorization and helps the students with active participation, to learn the subjects in their respective regional languages.

With the objective of applying ICT in Regional Language Education, in this paper we express the methodology of developing concept maps for various subjects at the higher secondary/pre-collegiate level.

2. Related Work

As it is known, an essential aspect in the learning process (either electronic or traditional) is the possibilities to evaluate the students. It is very important both for professor and student to test the understanding degree of the course. One of the best possibilities is to ask questions from the studied course. It is tested this way the degree of understanding of each studied material and the integration of new knowledge with the previous ones (that should already be known). These facts will have as a result an in-depth understanding of the learning materials.

Here we discuss the study and research done in connection with the proposed topic, by various experts outside India.

2.1 Question Generation for Learning Evaluation

Taking into consideration the high number of learning material existing in electronic format, the importance of the testing and evaluation systems has increased. The authors [McGough et. Al., 2008] present an interesting solution to the problem of presenting students with dynamically generated browser-based exams with significant engineering mathematics content. Here, the main idea is to generate the questions automatically based on question templates which are created by training on many medical articles. Liana et al [Liana Stanescu et. Al., 2008] tried to design and implement a software instrument (Test Creator) that permits generation of questions based on electronic materials that students have. The solution implies teachers to have a series of tags and templates that they have to manage. These tags can be used to generate questions automatically.

2.2 Concept Maps applied for Question Generation

Concept maps are a result of Novak and Gowin's (1984) research into human learning and knowledge construction. One of the powerful uses of concept maps is not only as a learning tool but also as an evaluation tool, thus encouraging students to use meaningful-mode learning patterns. Concept mapping may be used as a tool for understanding, collaborating, validating, and integrating curriculum content that is designed to develop specific competencies. Concept mapping, a tool originally developed to facilitate student learning by organizing key and supporting concepts into visual frameworks, can also facilitate communication among faculty and administrators about curricular structures, complex cognitive frameworks, and competency-based learning outcomes.

However, the only issue is that the learner must choose to learn meaningfully. The one condition over which the teacher or mentor has only indirect control is the motivation of students to choose to learn by attempting to incorporate new meanings into their prior knowledge, rather than simply memorizing concept definitions or propositional statements or computational procedures. The indirect control over this choice is primarily in instructional strategies used and the evaluation strategies used. Instructional strategies that emphasize relating new knowledge to the learner's existing knowledge foster meaningful learning. Evaluation strategies that encourage learners to relate ideas they possess with new ideas also encourage meaningful learning.

2.3 Concept Maps in E-learning

Recent researches have demonstrated the importance of concept map and its versatile applications especially in e-Learning. Concept maps creation for emerging new domains such as e-Learning is even more challenging due to its ongoing development nature. Concept maps can provide a useful reference for researchers, who are new to the e-Learning field, to study related issues, for teachers to design adaptive learning materials, and for learners to understand the whole picture of e-Learning domain knowledge.

2.4 Concept Map Mining

There is yet another approach [Villalon and Calvo, 2009] for automatic concept extraction, using grammatical parsers and Latent Semantic Analysis. Essays, as any other text, represent both the knowledge and the writing skills of its author; hence, an Automatic Concept Map from Essay (ACME) should reflect both. Therefore, the words for the concepts and relations must be extracted literally from the document, and the hierarchy of concepts must reflect the importance of the concepts relative to what was written in the particular document. However, the performance is related to the way concepts are chosen by humans. We believe that understanding this phenomenon and using it for the automatic selection of concepts could lead to big improvements.

3. Evaluation Mechanism

The proposed answer evaluation system creates concept-maps from the prose and the answer passages, that involves the extraction of concept-words from the passages by parsing the input text [Saravanan et al, 2004], creating a dependency model among the concept-words within the passage, visualization of concept-maps that has two levels viz., words level and the sentence pattern level that helps us to analyze what are the key concepts needed in a passage, Comparison of the concept-maps created and finally evaluation of the answer passages. The evaluation of the passage does not handle the content comparison part alone but also analyzes the finer aspects like the structuring

of the passage, repetitions, duplications, and also the additional aspects used by the students like causation, summarization, expansion at instance, etc. These finer aspects have been inspired from 'nannool'[12], a Tamil prose that presents the needed and irrelevant aspects of a prose.

For drawing light on the evaluation mechanisms followed, we choose a specific prose and answer and evaluate them. The total number of words in the prose above which was considered for the evaluation is 15. The comparison procedure followed is presented below.

Question

கம்பர் பற்றிய பெருமைகளை எழுதுக.

Prose

கல்வியில் பெரியவர் கம்பர். கம்பன் வீட்டுக் கட்டுத் தறியும் கவிபாடும். விருத்தமெனும் ஒன்பாவில் உயர்கம்பன். இத்தொடர்களால் கம்பருடைய பெருமைகளை அறியலாம்.

Answer

கம்பர் கல்வியில் பெரியவர். கம்பர் கல்வியில் பெரியவர். கம்பன் வீட்டுக் கட்டுத் தறியும் கவிபாடும். இத்தொடர்களால் கம்பருடைய பெருமைகளை அறியலாம். அவர் இராமாயணத்தை எழுதினார். அதனால் அவர் புகழ்பெற்றார்.

3.1. Missing Concepts

Comparing the prose and answer we find that the third sentence in prose is completely missing in answer. These three words will subtract 3 from 15. The last sentence in answer is not present in prose that's an irrelevant concept and we don't add any marks for it.

$$\text{Marks}=15-4 = 11$$

$$(11/15)*100 = 73.33\% (\text{word_match})$$

3.2.Duplicates

In answer the first two sentences are one and the same. We detect this with the help of the relation table. There are 59 relations generated and only 28 matches. So this module score is $(28/59) * 100 = 47.45\%$ (pattern).

3.3. Repetition

Sentence one is repeated again, so we deduct one mark. Therefore one mark should be deducted from the total score.

3.4. Value Addition

The evaluation method, value addition is computed based on causation and summarization, expansion at instance, argument by example. Here, the answer passage has summarization. Therefore one mark is to the total marks.

3.5. Score

From the values calculated above, the final score can be determined. Here, the given answer passage was evaluated to obtain the score 60.39% $((73.33 + 47.45)/2 + 1 - 1)$. Below the answer evaluation system results for the same prose and answer is provided.

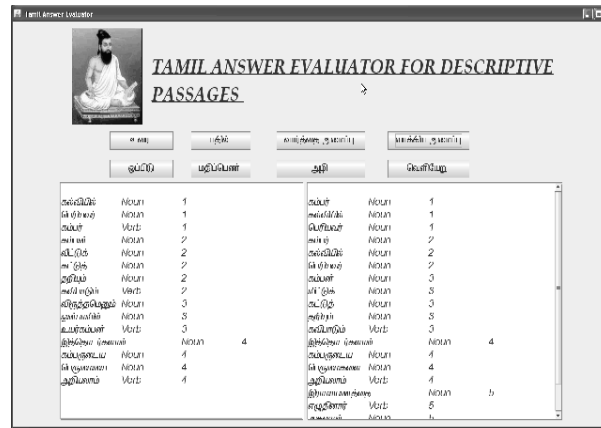


Figure 1 Screenshot for Answer Evaluation System

4. Conclusion

If teaching-learning educational process is considered as a goal through which students can get a meaningful learning of stated concepts, which extend and articulate their network of relations and can apply them in different contexts, it is necessary that teachers include tools to speed up act performance of agents involved in the construction of the new knowledge. In our case, applying a concept map tool in the classroom will allow students being themselves more motivated to carry out proposed activities and to participate in the construction of their own knowledge.

The methodology of developing concept maps discussed in the paper shall be (i) Extended to impart concept map based learning in other regional languages; (ii) Applied to generate associated concept animations for enriching automated content development; (iii) Applied to automated answer evaluation thereby taking part in self-assessment activity of pre-collegiate examinations; (iv) Used to dynamically generate questions and further continue the answer evaluation process in an e-learning setting; and (v) Applied for automatic document summarization

References

1. Jorge Villalon, and Rafael A. Calvo, "Concept Extraction from student essays, towards Concept Map Mining", Ninth IEEE International Conference on Advanced Learning Technologies, 2009. pp.221-225
2. Liana Stanescu, Cosmin Stoica Spahiu, Anca Ion, Andrei Spahiu, "Question generation for learning evaluation", Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 509 - 513, 2008 IEEE
3. McGough J., Mortensen J., Johnson J., Fadali S., "A web-based testing system with dynamic question generation". LNCS 1611-3349, 2008, pp. 242-251.
4. Novak J. and Gowin, Learning how to learn. New York and Cambridge, UK: Cambridge University Press, 1984.
5. R. Saravanan, Ranjani Parthasarathi and Geetha T.V., 'Vaanavil - Parser for Tamil', Resource Center for Indian Languages - Tamil, Dept. of Computer Science and Engineering, Anna University Chennai, India, 2004.
6. பவணந்தி முனிவர், "நன்னூல்", downloaded from <http://www.projectmadurai.org/pmworks.html>, accessed latest by April, 2010

A Machine Translation System for Converting Tamil Text -To -Sign Language

D.Narashiman

Student, M.E. Multimedia Technology-CEG
Anna University, Chennai-600 025
dnarashiman@gmail.com

Dr.T.Mala

Senior Lecturer
Anna University, Chennai-600 025
mala@cs.annauniv.edu

Abstract

This scheme enhances a remarkable approach to decipher the given Tamil text into sign language. The system suggested here is cross modal. It gives a wider interaction of textual input and generates corresponding sign language output in one or a number of sign variants. The application receives Tamil text sentences as input and provides output in 3D animated sequence that can be visualized.

Keywords-Sign language, 3D animation, machine translation.

Introduction

Sign language is a universal language and signs are common features in all the countries around the world. This system has been developed for the deaf community to enhance their communicative ability. This system was developed in such a way that it accepts text as input through keyboard(tactile mode) and generates animated sign symbols as output (visual mode).The benefits of this system include (a) maneuvered for many personal and real time confessions. (b) essential and vital tool for educative purpose to people who are interested in knowing the sign language.(c)The 3D animated output, endows the user to visualize the signs effectively for number of purpose such as news broadcasting and picture animation. This paper talks about the interactive system which translates Tamil text to sign language by compiling of sign language video [5], mapping of tokens and sign symbols [10][11] and creating animated output [3][12]. The sections are arranged as follows: Section two presents the literature survey on sign language. Section three describe the system architecture in general and also explain various phase in generating sign language in detail. Section four gives the method for evaluation of the translation output. Last section deals with conclusion and future enhancement of the system.

Literature Survey

Sign language is represented by non gesture, gesture coding. Non-gesture features include hand movements alone whereas gesture features include hand, face, shoulders, head and facial expression combined with hand sequences. Since sign variants are comparatively fewer than Tamil words the system should be able to generate sign for that corresponding word instantly. Sign language is lexically backward, it has become inventible for a sign generator to exhibit some degree of creativity in assigning concept-to-sign correspondence. When converting the Tamil text into sign language the creative method like finger spelling is used. The sign language mainly deals with non-gesture and gesture features of the signer. In practice however it is highly impossible to separate the context from the given text. Instead we prefer the semantic or conceptual representation of the source text. Various works were carried out earlier to solve this problem. Former approaches for converting text sign language are:

Example-Based method: Morrissey and way investigate corpus-based methods for example-based sign language translation from English to sign language of the Netherlands. With the small corpus and no available lexicon, the system is robust for sentences already encountered in the training set but has problems with unseen combinations of corpus chunks as well as corpus parts that it is unable to align [7].

Theoretical Methods: This method deals with the theoretical issues arising during machine translation of written text to sign language for example a notation for sign which use the 3D space around the signer to form complex expression [3] [4].

Rule-based Method: Safar and Marshall propose a decomposition of the translation process in two steps: initially they translate from written text into a semantic representation of the signs. Afterwards graphically oriented representation is done. Both the steps use rule-based techniques for a specific domain in British sign language [6].

Interlingua Method: An interlingua try to capture the generic fact – stating capacity of language using two different strategies: the first attempts to construct a universal grammar that generalize over the semantics nuances of many language, while the second is knowledge intensive allows for the incorporation of heterogeneous common sense into translation process [8].

Statistical Machine Translation method: This method is used to automatically transfer the meaning of the source language sentences into a target sentence by applying the phrase-based statistical machine translation based on morpho-syntactical analysis [1][2].

System Architecture

The system architecture design is given in figure 3.1,. The system mainly focuses on the construction of the direct knowledge repository and incremental knowledge repository. The finger spelling knowledge repository (F.S K.R), spatial knowledge repository and rule based knowledge repository comprise the incremental knowledge repository. System uses the knowledge repositories to translate given text to sign language. The pre-processor unit fed with parallel corpus. It converts the given input Tamil text and Sign videos into a suitable form for statistical analysis. The input text is divided into tokens using the white space as delimiter. The sign videos are also segmented. These tokens and corresponding signs are stored in a direct knowledge repository. The tokens for which signs are not available directly are generated and stored in incremental knowledge repository.

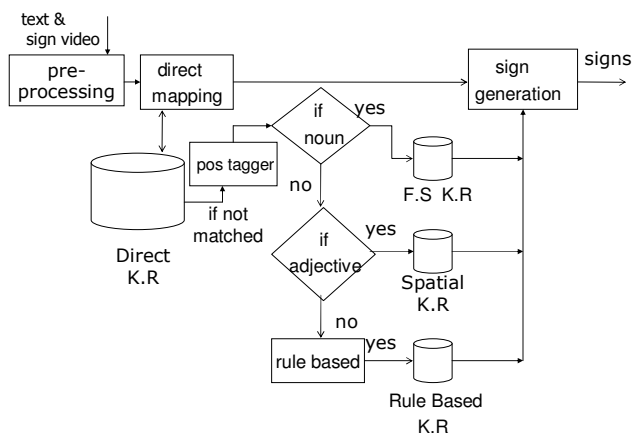


Figure 3.1 System Architecture

A. Creation of Direct Knowledge Repository

The creation of direct knowledge repository is shown in figure 3.2.

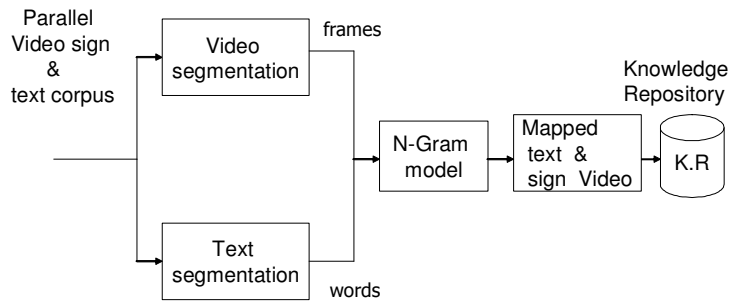


Figure 3.2 Creation of Direct knowledge Repository

Pre-processing is to convert the given input Tamil text and Sign videos into a suitable form. The text is processed by removing unimportant stop words, special characters and splits into words. Then the pre-processed words are stored in a table along with probability of occurrence using N-Gram methods. This is training the system. While training, statistical systems track common N-grams, learn which translations are most frequently used, and apply those meanings when finding the phrases in the future.

They also analyze the position of N-grams in relation to one another within sentences, as well as words grammatical forms, to determine correct syntax. The system uses the training to develop translation models. The system is fed by the parallel corpus. The language model is applied to the sign images. The system tracks bi-grams and builds the language model. The system then develops translation model from the parallel corpus and uses its training to translate new sentences.

The Direct Knowledge repository contains the index image in which segmented videos are indexed and stored in a database, alignment matrix is constructed in order to obtain the matching of word-to-sign coherence, and Language Model Probability. The language model is applied to the target language (sign language) for reordering and arrangement of the signs. N-Gram calculations are applied and probability of the sign occurring is calculated and stored in the table. Where probability of the output sign depends on the given text and probability of that sign generated for the given text.

A. Creation of incremental Knowledge Repository

The word which does not inherit appropriate sign by using direct knowledge repository is exhibited using incremental knowledge repository. The incremental knowledge repository is created by finger spelling, spatial and rule based method as shown in the figure 3.3. The purpose of this to generate sign for the words not present in the direct knowledge base. The proper noun words, descriptive word and ambiguous words come under this category.

The parts of speech tagger is used to group the words according to proper noun, adjective, ambiguous and action words. Depending upon the classification the sign are generated from respective incremental knowledge repository. The word whose sign are not generated by direct knowledge repository are generated by incremental knowledge repository. The preprocessed word is tagged using the pos tagger

which gives the information about the word whether proper noun, action and descriptive words according to the classification the process is carried out.

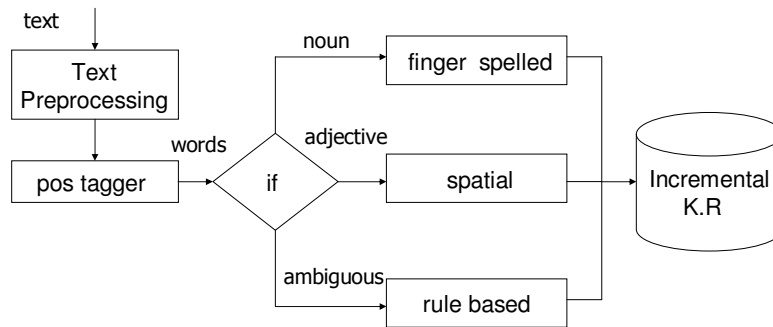


Figure 3.3 Creation of Incremental knowledge Repository

The proper noun words are represented using this knowledge repository. The process is shown in figure 3.4. This repository contains sign for individual alphabet. The word is tagged by POS tagger if the word is proper noun it is again rechecked using the word net dictionary if synonyms is available it is signed using direct knowledge repository otherwise the word is separated into individual characters and signed respectively. Eg: பெயர்ச்சால்- ெௌௌௌ. The name is separated into signal characters and they are signed.

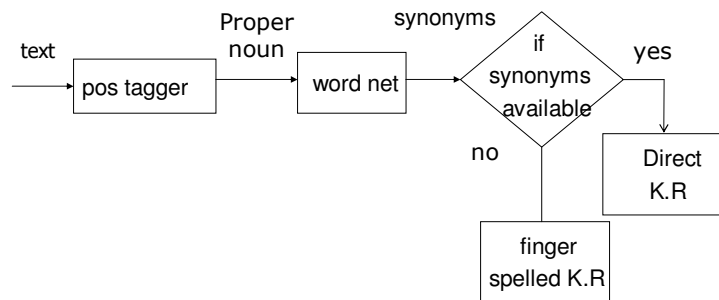
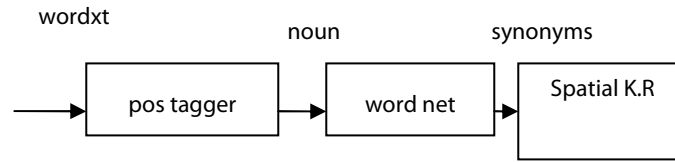


Figure 3.4 Finger Spelling method

Spatial method is used to represent the descriptive words. The method is shown in figure 3.5. Usually these words describe the place, nature, behavior etc. The main purpose of this module is to reduce the number of sign variants as the certain words have common meanings and the same sign is represented for different words. This process begins with the tagging of the word and check whether the word is descriptive (adjectives), the meaning for the word is retrieved using the word net dictionary then the meanings are compared with the words in the spatial knowledge repository if the meaning match the word then corresponding sign is generated. By this method the signs for different words are generated easily with minimum number of signs video in the database. Eg: ூௌௌ , ூௌௌ both the words represent the same meaning of spacious , the same sign is represented for both the word.

Figure 3.5 Spatial method



Rule based method is used represent the ambiguous word. The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word eg : $\text{\textcircled{A}}$ - it may represent either a day of the week or sun depending on the context the sign is generated. A word is assumed to have a finite number of discrete senses often given by a dictionary and task of the forecast is to make a forced choice between these senses for the meaning of each usage of an ambiguous word based on the context of use. This procedure is termed as context based disambiguation. The word is tagged if the tag is ambiguous then the pervious word tag is considered depending on previous word the sign is generated. The alternative method is using n-gram model. An n-gram is a type of phrase within a document that has a set of n number of words. N-grams are the basic linguistic unit with which a statistical machine translation system works. The n-gram gives the maximum probability of word occurring together. This is used to check the word occurrence in the context and along with the other words in sequence. This also used to determine the word that are unable to decode correctly, by checking the context and word occurrence. From this the sign for ambiguous word are generated The rule based method is shown in the figure 3.6.

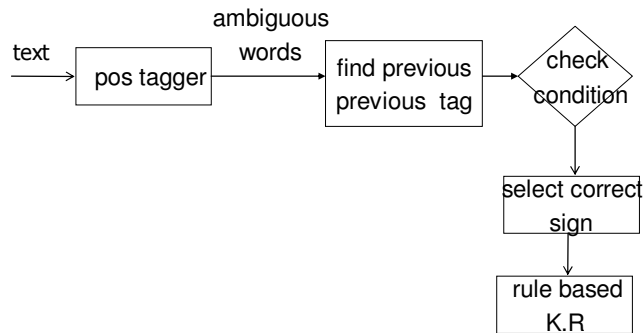
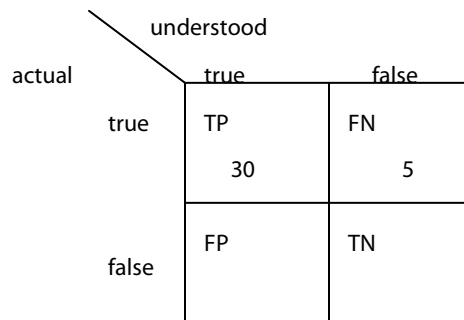


Figure 3.6 Rule based method

Finally signs are arranged in the logical sequences and the final output is generated. The final output is in the form of 3D animation. If sign for corresponding word not found in the direct knowledge repository then the word is tagged using pos tagger. According to the tag, if the word is proper noun then the signs are generated using finger spelling knowledge repository, if the word is an adjective then spatial knowledge repository is used and if the word are ambiguous the rule based knowledge repository is used.

Evaluation Criteria

The output of the system is final as it should generate the correct sign for the given text. The generated output should be clear, understandable and correct. The hand and facial movements should clear and various actions are understandable. The correct matching of sign and the word should be done. The output of the system is validated using the confusion matrix. The confusion matrix is plotted by taking into consideration the actual generated sign and how it understood. The first quadrant represent the correct sign and understood correctly, second quadrant represent the correct sign but understood incorrectly, third quadrant represent incorrect sign generated but understood correctly and the last quadrant represent the incorrect sign generated and understood incorrectly.



The diagram shows a 2x2 confusion matrix. The vertical axis is labeled 'actual' with 'true' and 'false' categories. The horizontal axis is labeled 'understood' with 'true' and 'false' categories. The cells contain: TP (30) for (actual true, understood true), FN (5) for (actual true, understood false), FP for (actual false, understood true), and TN for (actual false, understood false).

		understood	
		true	false
actual	true	TP 30	FN 5
	false	FP	TN

Figure 4.1 Confusion matrix

The confusion matrix shown in the figure 4.1 about 50 words are given as input and output signs are checked, The value in the first quadrant 30 are the signs generated are correct and understood correctly. The value in the second quadrant is obtained when there is difference in retrieval of sign number from the database. The value in the third quadrant is obtain as there is difference in generating spatial sign. The fourth quadrant value is obtain when there is difference in retrieval of sign for the given word as the word is an ambiguous word.

Conclusion

The system described here is very useful for the deaf community. The process starts by breaking the sentence into tokens, mapping the tokens with the sign symbols, generating the perfect symbol in accordance with the conditions and rearranging of symbols to generate the correct meaning of the sentence. The output is a real visualization of sign language and enables interactive application. The main challenge of the system is mapping the sign , generating the new sign if not available in the database and rearranging the symbols according to the meaning of the text checking the grammar of the sentence , adding to the database for future reference. As it is not the only way to generated sign symbols it can be always impoverished to the native speakers. The animation output is selected as it will be pleasing for the viewers. This is mainly designed as a non-human interpreter for sign language. Future development in this system is to increase the sign symbols for the words that are commonly used in day to day life and also generalizing the symbols irrespective of native speakers, to customize the output image according to their interest. The system can be extended to any language by using a language translator with little modification and it also an be extended to translate speech to text and then to sign language.

References

- [1] Daniel Stein, Jan Bungeroth, and Herman Ney, "Morpho-syntax based statistical methods for automatic sign language translation," in proceedings of European Association for Machine Translation, 2006, vol 10, pp.169-177.
- [2] S.Kanthak, D.Vilar, E.matusov, and H.Ney, "Novel rendering approaches in phrase-based statistical machine translation," in proceedings of Association of Computer Linguistics Workshop, 2005, pp.167-174.
- [3] Maria Papadogiorgaki, Nikos Grammalidis, Dimitrios Tzovaras, and Michael G.Strintzis, "Text-to-sign language synthesis tool," in proceedings of European Signal Processing Conference, 2005, pp.130-134.
- [4] Matt Huenerfauth, "Spatial and planning models of ASL classifier for machine translation," in proceedings of Theoretical and Methodological Issues in Machine Translation, 2004, pp.65-77.
- [5] Mohamed Mohandes, "Automatic translation of arabic text to arabic sign language", Artificial Intelligence and Machine Learning, 2006, vol.6(4), pp.15-19.
- [6] E.Safar and I.Marshall, "The architecture of an English text -to-sign languages translation system," in proceedings of Recent Advance in Natural Language Processing, 2001, pp.223-228.
- [7] Sara Morrissey and Andy Way, "An example-based approach to translating sign language," in proceedings of Example Based Machine Translation, 2005, vol.11, pp.106-116.
- [8] Tony Veale, Brona collins, and Alan Conway, "The challenge of cross -modal translation proceedings of Association of Machine Translation, 2000, vol.13(1), pp.81-106.
- [9] R.Zens, O.Bender, S.Hasan, S.Khadivi, E.Matusov, J.Xu, Y.Zhang, and H.Ney, "The RWTH phrase-based statistical machine translation system," in proceedings of International Workshop on Spoken Language Translation, Oct 2005.
- [10] R.Zens, F.J.Och, and H.Ney, "Phrase -based statistical machine translation," in proceedings of German conference in Artificial Intelligence, 2002, vol.2479, pp.18-32.
- [11] Yuging Guo, Josef van Genabith, and Haifeng Wang, "Dependency -Based Ngram models for general purpose sentences realisation," in proceedings of International Conference on Computational Linguistics '08, 2008, pp.297-304.
- [12] D.Narashiman and T.Mala, "An Interactive System for Converting Text-to-Sign Language," in proceeding of International Conference on Information Systems & Software Engineering, 2009, pp.164-167.

இயந்திர மொழிபெயர்ப்பிற்கான அகராதி

உருவாக்கம்

Lexicon for Machine Translation

முனைவர் ப. குமார்

கௌரவ விரிவுரையாளர்

தமிழ் மொழித்துறை

சென்னைப் பல்கலைக்கழகம்

kumarchomsky@gmail.com

அச்ச அகராதியும் மனித மூளை அகராதியும் (Printed Lexicon and Mental Lexicon)

ஒருமொழி அகராதி, இருமொழி அகராதி, பன்மொழி அகராதி, வட்டாரச்சொல் அகராதி, பேரகராதி, படவிளக்க அகராதி, தற்காலத் தமிழ் அகராதி என அகராதிகளில் பலவகையுண்டு. துறைசார் அகராதி வகையில் அறிவியல் அகராதி, ஆட்சிச்சொல் அகராதி, கலைச்சொல் அகராதி எனவும் பலவகைப் படுவதுண்டு. இவையனைத்தும் இயற்கை மொழியறிவைக் (Knowledge of Natural Language) கொண்ட மனித மூளைக்கான அகராதிகளாகும்.

அச்ச அகராதிகளில் ஒரு சொல்லிற்கு, மொழிச்சூழலின் பாதிப்பு இன்றி அதற்கான பொருளை மட்டுமே காணமுடியும். ஆனால், அச்சொல் தொடர்நிலையில் பயின்றுவரும்போது மொழிச்சூழலுக்குட்பட்டு எவ்வாறு வேறுபட்ட பல்வேறு பொருள்களைத் தருகிறது என்பதையும் ஒவ்வொரு பொருளும் எந்தெந்த அடிப்படையில் வேறுபடுகிறது என்பதையும் அச்ச அகராதிவழி அறியவியலாது.

‘ஓடு’ என்னும் வினைச்சொல்லை எடுத்துக்கொண்டால் அது கீழ்வரும் பொருள்களைத் தருவதைக் காணமுடியும்.

அவன் வேகமாக ஓடினான்.

“He runs fast”

புதிதாக வந்திருக்கும் இந்தப் பொருள் எங்கள் கடையில் சுமாராக ஓடுகிறது.

“This thing which arrived recently in our shop, moves very slowly.”

அவர் திடீரென்று வந்தவுடன் எனக்கு ஒன்றுமே ஓடவில்லை.

“As he comes suddenly I become functionless”

‘ஓடு’ என்னும் சொல்லிற்கு இணையாக ஆங்கிலத்தில் ‘run’, ‘flow’ போன்ற சொற்கள் இருக்கின்றன. எந்த இடத்தில் ‘run’, என்று மொழிபெயர்க்கவேண்டும், எந்த இடத்தில் flow’ என்று மொழிபெயர்க்க வேண்டும் என்பதைக் கணினி எப்படிப் புரிந்து கொள்ளும்?

ஒரு சொல் தனிநிலையில் வரும்போதும், ஒரு சொல்லிற்குப் பல பொருள்கள் இருக்கும்போதும் ஏற்படும் பொருண்மைச் சிக்கல்கள் அல்லது மயக்கம் (Sense ambiguity) இவ்வகை அகராதி உருவாக்கத்தில்

கவனிக்கப்பட வேண்டிய ஒன்றாகும். ஒரு சொல் தொடர் நிலையில் வரும்போது, அச்சொல் முந்தைய, பிந்தைய சொல்லின் பண்பிற்கு ஏற்பவும் பேசப்படும் சூழல், காலம், உரையாடும் நபர்கள் போன்ற பிற மொழிசாராக் கூறுகளைப் பொறுத்தும் தன் பொருண்மையை எவ்வாறு மாற்றிக்கொள்கிறது என்பதையும் கவனத்தில் கொள்ளவேண்டும்.

எடுத்துக்காட்டாக,

1. 'பச்சைத் தண்ணீர்' என்பதிலுள்ள 'பச்சை' என்னும் சொல் 'தண்ணீர்' என்னும் சொல்லின் பண்புகளை ஏற்று சூடாக்கப்படாத அல்லது குளிர்ந்த என்னும் பொருளில் வருகிறது.
2. 'பச்சைக் குழந்தை' என்பதிலுள்ள 'பச்சை' என்னும் சொல் குழந்தையின் பண்புகளை ஏற்று 'பிறந்த' அல்லது 'சிறுவயது குழந்தை' என்னும் பொருளில் வருகிறது.
3. 'பச்சை சட்டை' என்பதிலுள்ள 'பச்சை' என்னும் சொல் சட்டையின் 'நிறப் பண்பை' ஏற்று வருவதைக் காணமுடிகிறது.

'பச்சை' என்னும் சொல்லிற்கு வெவ்வேறு சொற்பொருண்மைகள் (word senses) இருந்தாலும் அவை யாவும் எவ்வாறு வேறுபடுகின்றன என்பதையும், பல்வேறு சூழல்களில் வரும்போது அதற்கேற்ற பொருண்மையாக மாறிவிடுகின்றன என்பதையும் இயற்கை மொழியறிவைக் கொண்ட மனித மூளை நன்றாக அறியும். மனித மூளையானது அச்ச அகராதியில் ஒரு சொல்லுக்குக் கொடுக்கப்பட்ட பல்வேறு சொற்பொருண்மைகளில் குறிப்பிட்ட தொடருக்கு ஏற்ற ஒரு பொருண்மையை எடுத்துக்கொண்டு, அதை மூளை அகராதியில் (mental lexicon) அச்சொல்லுக்குக் காணப்படும் பிற பொருண்மைப்பண்புகளோடு இணைத்து , பொருளைக் (meaning) கண்டறியும் அல்லது முடிவு செய்யும்.

அச்ச அகராதிகளில் உள்ள ஒரு சொல்லிற்கான பொருளை நம் மூளை எளிதாகப் புரிந்து கொள்கிறது. அதற்கு அடிப்படைக் காரணம், நம் மூளையில் உள்ள மூளை அகராதியாகும் (mental lexicon). நாம் புற உலகில் ஒரு சொல்லைக் கேட்கும்போது, அச்சொல்லை நமது மூளை அகராதியின் உதவியுடன் புரிந்து கொள்கிறோம். மூளை அகராதியில் சொற்களுக்குக் கொடுக்கப்பட்டுள்ள பல்வேறு பொருண்மைக் கூறுகளில் ஒரு சிலவற்றை அச்ச அகராதியில் நாம் காண்கிறோம். அச்ச அகராதியில் கொடுக்கப் பட்டுள்ள பொருண்மைக்கூறுகளை அடிப்படையாகக்கொண்டு, நாம் மூளை அகராதியின் உதவியுடன் சொற்களின் பொருளைப் புரிந்துகொள்கிறோம். மேலும் மூளையின் மொழிப்புலனானது மூளை அகராதி அளிக்கும் தகவல்களைப் புற உலகத்தைப்பற்றிய பிற அறிவுகளையும் பின்புலனாகக்கொண்டு, ஒரு தொடரின் பொருளைப் புரிந்துகொள்கிறது.

பொதுவாக, மொழியில் பொருள் தெரியாதபோதுதான் நாம் அச்ச அகராதிகளைப் பயன்படுத்துகிறோம். அவ்வாறு பொருள் தெரியாத நிலையில் அகராதிகளில் உள்ள ஒருசில கூறுகளுடன் மூளை அகராதியிலுள்ள பல்வேறு கூறுகளையும், அப்பொருள் சார்ந்த புறவுலக் கூறுகளையும் கொண்டுதான் முழுப்பொருளையும் புரிந்துகொள்கிறோம்.

'படி' என்னும் சொல்லிற்கான இலக்கணப் பொருளை அச்ச அகராதியில் காணும்போது அது பெயர்ச் சொல்லா, வினைச்சொல்லா என்பதை முடிவு செய்வதற்கு அச்ச அகராதியிலுள்ள இலக்கணக்கூறு நமக்கு வெளிப்படுத்தும்.

'அனுமதி' என்னும் சொல்லிற்கான சொற்பொருண்மையை அச்ச அகராதியில் காணும்போது பல பொருண்மைகள் இருப்பதை உணரமுடிகிறது. குறிப்பாக, 'அவர் மருத்துவமனையில் அனுமதிக்கப் பட்டார்', என்னும் தொடரில் 'அனுமதி' என்னும் சொல் தரும் பொருளும், 'கண்ணன் அலுவலகத்திற்குச் செல்ல அனுமதிக்கப்பட்டான்' என்னும் தொடரில் 'அனுமதி' என்னும் சொல் தரும் பொருளும் வெவ்வேறாக உள்ளது.

இவ்வேறுபாட்டை அச்ச அகராதியில் காணவியலாது. இவ்வேறுபாட்டை நமக்கு உணர்த்துவது அச்ச அகராதிக் கூறுகள் மட்டுமல்லாது அதையும் தாண்டியதொன்று செயல்படுவதை நன்கு உணரமுடிகிறது. இதைத்தான் மூளை அகராதி என்கிறோம். அச்ச அகராதியில் ஒரு சொல்லிற்கான பொருள் விளக்க சில கூறுகள் இருந்தாலும், இதனுடைய பிற கூறுகள் மூளை அகராதியில் இருந்து எடுத்துக்கொண்டு அதனுடன் புறவுலகப் பொருண்மைகளையும் இணைத்துக் கொண்டுதான் நாம் இப்பொருண்மையை உணருகிறோம்.

கணினியும் அச்ச அகராதியும்

கணினியானது இயற்கைமொழிச் சொற்களின் பொருளை எவ்வாறு புரிந்துகொள்கிறது? அச்ச அகராதி தரும் உதவியைப் பெற்றுக்கொண்டு மனிதமூளையானது தனது மூளை அகராதியின் துணையுடன் சொற்பொருளைப் புரிந்துகொள்வதுபோல, கணினியால் புரிந்துகொள்ளமுடியுமா? மூளை அகராதிக்கு இணையான ஒரு அகராதி கணினிக்கு இருந்தால், அதனால் மூளைபோன்று சொற்பொருளைப் புரிந்துகொள்ளமுடியும். அச்ச அகராதியைப் பயன்படுத்தமுடியும்.

அப்படியென்றால், கணினிக்கு இயற்கைமொழிச் சொற்களின் பொருள்களைப் புரிந்துகொள்ள ஒரு தனிப்பட்ட அகராதி - மூளை அகராதியை ஒத்த ஒரு அகராதி - உருவாக்கப்படவேண்டும். இதுபற்றிய ஆய்வில் ஈடுபட்டுள்ள கணினிமொழியியலர்கள் கணினி அகராதிக்கான பல்வேறு கோட்பாடுகளையும் வடிவங்களையும் முன்வைத்துள்ளார்கள். மில்லர் என்பவர் சொல்வலை (WordNet) என்ற ஒன்றையும் பெஸ்டோவெட்ஸ்கி என்பவர் உருவாக்க அகராதி (Generative Lexicon) என்ற ஒன்றையும் முன்வைத்துள்ளனர்.

ஒரு குறிப்பிட்ட மொழிக்கான கணினி அகராதியில் மூளை அகராதியில் உள்ள அனைத்துக்கூறுகளும் இடம்பெறுமாறு செய்யவேண்டும். அதனுடன் மனிதமூளைக்கு உள்ள உலகியல் அறிவு இல்லாததால், அப்பற்றாக்குறையையும் ஈடுகட்டக்கூட வகையில் அது அமையவேண்டும்.

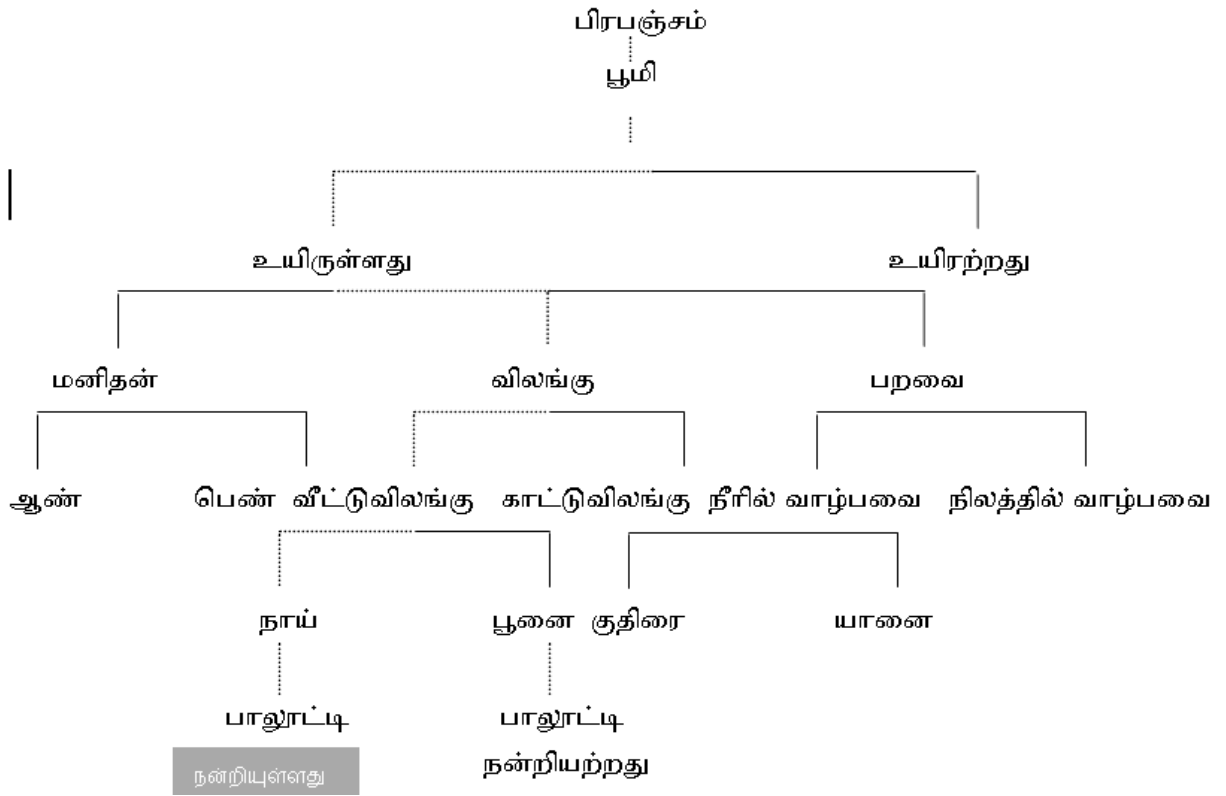
மனித மூளையில் உள்ள அகராதி அறிவை எவ்வாறு கணினியில் முன்வைப்பது (knowledge representation) என்பதே அடிப்படை வினா. இதற்கான விடையைக் கணினிமொழியியல் தருகிறது. மேற்குறிப்பிட்ட அகராதி அறிவை தரவுத்தளமாகவும் (database structure) நிரல்வழிமுறைகளாகவும் (algorithms) மாற்றியமைத்து, கணினிக்குக் கொடுக்கப்படுகிறது. மில்லரின் சொல்வலையும் பெஸ்டோ வெட்ஸ்கியின் உருவாக்க அகராதியும் இதற்கான வழிமுறைகளை முன்வைக்கிறது.

சொல்வலை (Wordnet)

கணினிக்கான அகராதி உருவாக்கம் என்பது பொருண்மை அடிப்படையிலானது. ஒரு சொல்லிற்கான உறவுநிலைப் பொருண்மையியல் (Relational Semantics), மற்றும் சொற்சேர்க்கைகளைக் கண்டறிவது இயந்திர மொழிபெயர்ப்பிற்கு முக்கியமானதாகும். சொற்பொருள்களுக்கான உறவுநிலைகள் என்பது ஒரு பொருள் தன்னோடு தொடர்புடைய மற்றொரு பொருளுடன் எவ்வாறு உறவு கொண்டுள்ளது என்பதைக் கண்டறிவதாகும். 'ஒடு' என்னும் சொல் 'கூரை ஒடு', 'பாளை ஒடு' என்னும் இருபொருளைத் தருவதாகக் கொண்டால், 'கூரை ஒடு', 'பாளை ஒடு' இவ்விரண்டிற்குமான உறவை இனம் காண வேண்டும். 'கூரைக்கும் ஒட்டிற்கும்' உள்ள உறவு ஒரு முழுமை சார்ந்தது. 'பாளைக்கும் ஒட்டிற்கும்' உள்ள உறவு ஒரு பகுதி சார்ந்தது. ஒன்றின் முழுமைக்கும் பகுதிக்குமான உறவு என்ன என்பதைக் கண்டறிதல் முக்கியமானதாகும். இம்முறையைத்தான் அகராதிப் பொருண்மை (Lexical Semantics) என்கிறோம். ஒரு சொல்லின் பொருண்மை பற்றிய எந்தவொரு குறைந்தபட்சப் பகுப்பாய்வையும் உருவாக்கச் செயற்பாடு (Generative operation) எனக் கூறலாம். ஒரு சொல்லிற்கும் அதன் மூலக்கூறின் பகுதிக்கும் இடையே உள்ள உறவு, ஒரு சொல்லை ஒரு பெரும் பகுதிக்குள் வேறுபடுத்திக்காட்டுவதாகும். இதனைக் கீழ்வரும் சொல்வலை அமைப்பு நிறுவுகிறது.

நன்றியுள்ள விலங்கு எது? என்ற கேள்வி எழுப்புவதனுடாக பல்வேறு பதில்களை இவற்றிலிருந்து பெற்றுக்கொள்ளமுடியும். இந்த பிரபஞ்சத்திலுள்ள பூமியில் இருக்கும் உயிருள்ளவற்றில் விலங்கு வகையைச் சேர்ந்தது என்றும், அதிலும் வீட்டில் வளர்க்கப்படும் விலங்கினத்தைச் சேர்ந்ததுமான நாய் நன்றியுள்ளது ஆகும் என்ற பதில், மூலப்பொருளுக்கும் ஒரு சொல்லுக்குமான உறவை வெளிப்படுத்து கிறது.

‘இராமன் தசரதனின் மகன்’ என்ற தொடர் ‘இராமனுடைய அப்பா தசரதன்’ என்ற மற்றொரு பொருளையும் தருகிறது. ஒரு தொடர் தரும் பொருண்மையை மட்டுமல்லாது அவற்றினூடாக தர்க்க ரீதியான (Logic) பொருளையும் உருவாக்க இயல்வது மனிதமுனை. இவ்வாறு தர்க்கரீதியான சிந்தனை என்பது கணினியில் கொண்டு வர இயலாது. இதுபோன்று கணினியும் ஒரு பொருளைப் புரிந்து கொள்ள வேண்டுமெனில் ஒரு சொல்லிற்குரிய பல்வேறு குறிப்புகளைக் (hinds) கொடுப்பதுதான் சொல்வலை ஆகும்.



உருவாக்க அகராதி (Generative Lexicon)

“உருவாக்க அகராதி (Generative Lexicon) என்பது ஒரு சொல்லிற்கான பல்வேறு பண்புகளைப் பொருண்மையடிப்படையில் கொண்டதும் இயந்திர மொழிபெயர்ப்பிற்குப் பயன்படும் பல்வேறு அகராதி வடிவமைப்புகளில் ஒன்றானதுமாகும்” (Pustejovsky J.). உருவாக்க அகராதி என்பது கிட்டத்தட்ட இயற்கை மொழியைக் கணினிக்குப் புரிய வைப்பதற்கான முயற்சியாகும்.

ஒருமொழி, இருமொழி, பன்மொழி வடிவிலான அகராதிகளைப் பொருத்தமட்டில், அதற்கான தலைப்புச் சொல், இலக்கணக் குறிப்பு, பொருள், அச்சொல் மொழியில் பயின்றுவரும் நிலை ஆகியவைக் காணப்படும். உருவாக்க அகராதிக்கு இத்தரவுகளோடு ஒருசொல்லின் பல்வேறு பண்புகளைப் பொருண்மையடிப்படையில் கொடுத்தாக வேண்டும்.

ஒரு வினைச்சொல் எடுக்கும் பெயர்நிலைகள் (Argument) எத்தனை என்பதும் கொடுக்கப்பட வேண்டும். 'கொடு' என்னும் வினைச்சொல் "கொடுப்பவர், வாங்குபவர், கொடுக்கப்படும் பொருள்" ஆகிய மூன்று பெயர்நிலைகளை ஏற்கும். இதுபோன்று ஒவ்வொரு வினைச்சொல்லும் எத்தனை பெயர்நிலைகளை எடுக்கும் என்பதும் இவ்வகை அகராதி உருவாக்கத்திற்கு முக்கியம்.

'நல்ல' (Good) என்னும் பெயரடை பெயருடன் இணைந்து 'நல்ல பையன்' (Good Boy), 'நல்ல தண்ணீர்' (Pure water), 'நல்ல நாள்' (Auspicious day), 'நல்ல வேளை' (Auspicious time), 'நல்ல நண்பன்' (Thickest/Best friend), 'நல்ல பாம்பு' (Cobra), எனப் பல நிலைகளில் வரும். 'நல்ல' எனும் சொல்லிற்கானப் பொருளை தீர்மானிப்பது அதனை அடுத்துவரும் பெயராகும். 'நல்ல பையன்' என்பதிலுள்ள 'நல்ல' எனும் சொல் 'ஒழுக்கம்' என்னும் பொருளையும், 'நல்ல தண்ணீர்' என்பதிலுள்ள 'நல்ல' எனும் சொல் 'சுத்தம்' எனும் பொருளையும், 'நல்ல நாள்' என்பதிலுள்ள 'நல்ல' எனும் சொல் 'ஏற்ற/ பொருத்தமான' என்னும் பொருளையும் பொருண்மை அடிப்படையில் தருவதைக் காணமுடிகிறது. இதுபோன்று ஒவ்வொரு சொல்லும் தொடரில் பயன்படுத்தப்படும்போதும் பல பொருண்மை அடிப்படை யிலானப் பண்புகளைக் கொண்டுள்ளன.

இந்நிலையில் 'உருவாக்க அகராதி'யில் 'நல்ல' எனும் சொல்லிற்கு இணையாக ஆங்கிலத்தில் 'Good' என்னும் சொல்லை மட்டும் கொடுத்தால் இதில் காணப்படும் பல்வேறு பொருண்மை அடிப்படையிலான கருத்துகளைப் புரிந்து கொள்ள முடியாமல் போய்விடும். 'நல்ல' என்னும் சொல்லிற்கு 'Good' எனும் சொல்லை மட்டும் பயன்படுத்தி மொழிபெயர்த்தால் 'நல்ல பையன்' என்பதை 'Good Boy' என்று மொழி பெயர்க்கலாம். ஆனால், 'நல்ல தண்ணீர்' என்பதை 'Good Water' என்று மொழிபெயர்க்க முடியாது. 'Pure water /Drinking water/Portable water' என்றுதான் மொழிபெயர்க்க முடியும். 'நல்ல' எனும் சொல் 'தண்ணீர்' எனும் சொல்லுடன் வரும்போது அதன் பொருண்மை அடிப்படையிலான கருத்தை நாம் புரிந்திருப்பதுபோல் கணினியும் புரிந்து கொள்ள வேண்டுமெனில், 'தண்ணீர்' பற்றிய பொருண்மை அடிப்படையிலான பண்புகளைக் [திரவமானது; குடிக்கக் கூடியது; எளிதில் மாசுபடக் கூடியது] கொடுக்க வேண்டும். அப்பொழுதான் கணினியும் 'நல்ல தண்ணீர்' என்பதிலுள்ள 'நல்ல' என்பதற்கும், 'நல்ல பையன்' என்பதிலுள்ள 'நல்ல' என்பதற்குமான பொருண்மை அடிப்படையிலான பொருளைப் புரிந்து கொள்ளமுடியும்.

உருவாக்க அகராதிக் கூறுகள்

கணினிச் சொற்பொருண்மையிலுக்கு எத்தகைய பொருண்மை வெளிப்பாட்டுப் படிநிலைகள் தேவை? என்கிற கேள்வி, ஒரு சொல்லை ஒரு தொடரில் எப்படிப் பொருத்துவது என்பதற்கு விடையளிக்கும். இதனை உருவாக்க அகராதிக் கோட்பாட்டை அறிவதன்மூலம் புரிந்து கொள்ள முடியும். இது நான்கு நிலைகளைக் கொண்டுள்ளது.

1. பெயர்நிலையமைப்பு (Argument Structure)
2. நிகழ்வமைப்பு (Event Structure)
3. பண்பமைப்பமைப்பு (Qualia Structure)
4. மரபமைப்பு (Lexical Inheritance)

இத்தகைய நான்கு அமைப்புகளும் சொற்பொருண்மையிலின் ஒரு கணினிக் கோட்பாட்டிற்குத் தேவைப்படுகிற பல்வேறு பொருண்மை வெளிப்பாடு மற்றும் வெளிப்பாட்டு நிலைகளை முக்கியமாகக் கொண்டுள்ளன. ஒவ்வொரு அமைப்பும் ஒரு சொல்லின் பொருண்மைக்கு வேறொரு விதமான தகவலைக் கொடுக்கும்.

இவ்வாறாக, மொழியை மனித மூளை புரிந்து கொள்வது பொருளை மட்டுமல்ல, அதனோடு சேர்ந்த மொழிச்சூழலையும் தான். அவற்றிலிருந்து ஒரு சொல்லிற்குச் சரியான பொருளைத் தேர்ந்தெடுப்பதற்கு மனித மூளைக்குச் சில செயல்கள் மொழிக்கருவிகள்போல இருந்து செயல்படுகின்றன என்பதை இதன் வழி அறிய முடிகிறது. ஆகவே கணினியின் மூலம் முழுமையான மொழிபெயர்ப்பினை அச்ச அகராதி களை மட்டும் கொண்டு செய்யமுடியாது என்பதும், அச்ச அகராதிக்கூறுகளுடன் சேர்ந்து பொருளைப் புரிந்து கொள்ளப் பயன்படும் மூளைஅகராதியின் கூறுகளையெல்லாம் புரிந்தால் மட்டுமே கணினி மொழிபெயர்ப்பிற்கான அகராதியை உருவாக்க இயலும் என்பதும் இதன்வழி கண்டறியப்பட்டுள்ளது.

பார்வை நூல்கள்

- பா.ரா. சுப்பிரமணியன். (பதிப்பாசிரியர்), சொல் வழக்குக் கையேடு, மொழி அறக்கட்டளை, சென்னை, 2005.
- பா.ரா. சுப்பிரமணியன். (பதிப்பாசிரியர்), க்ரியாவின் தற்காலத் தமிழ் அகராதி, க்ரியா பப்ளிகேஷன்ஸ், சென்னை, 2004.
- பா.ரா. சுப்பிரமணியன். மொழியில் மரபு வழிப்பட்ட சொற்சேர்க்கைகள் (கட்டுரை) சென்னை, 2002.
- கு. சுப்பையாப்பிள்ளை, இயற்கை மொழியாய்வு தமிழ், உலகத் தமிழாராய்ச்சி நிறுவனம், சென்னை, 2003.
- தமிழ்ப் பேரகராதி, சென்னைப் பல்கலைக்கழகம்.
- Pustejovsky, J., *The Generative Lexicon*, MIT Press, 1996.
- T. Burrow, M.B. Emeneau, *A Dravidian Etymological Dictionary*, Munshiram Manoharlal Publisher Pvt. Ltd., New Delhi - 1998.

Tamil Hyper Grammar

Uma Maheshwar Rao G

guraohyd@gmail.com

Christopher M

efthachris@gmail.com

Parameswari K

parameshkrishnaa@gmail.com

Center for Applied Linguistics and Translation Studies

University of Hyderabad, Hyderabad – 500046.

Abstract

Grammatical descriptions of human languages are the results of efforts in modeling of the design features and the internal organization of the structures and the mechanisms. Therefore, Linguistics is about language modeling, designing and studying their theoretical and practical implications. However the activity of grammatical descriptions itself is molded by the specific needs of aims and the goals such as Teaching and Learning a language, investigating the issues related to the evolutionary biology with regard to discovering the universals of human language and development, philosophical and functional aspects of language and Linguistic Computing. Here, we would like to discuss certain issues towards building a Hyper grammar for a given language.

1. Concept

A Hyper grammar is a non-linearly organized dynamic grammar based on the hypertext format. It is intended to simulate certain functions of a native speaker. It can be used both as learning and teaching tool besides as a reference grammar.

It is comprised of a number of non-linearly arranged texts each with a comprehensive note on various grammatical facts of Tamil, with hyper-links. It can be accessed and retrieved for various purposes involving language, to experience the effect of a native speaker and hearer of the language. Functionally it serves better than any of the existing printed grammars, which are simply flat and linear. In a way the existing printed grammars are non-communicative i.e. passive, hence, they are monologues and do not participate or reciprocate to pass judgments about the linguistic facts of the respective languages. A grammar in order to reciprocate should have some of the computationally implemented tools like a morphological generator, analyzer, chunker, parser, lexical accessor etc.

The Hyper grammar is intended to be a reciprocative grammar, as it involves some of the properties like the native speaker's ability to make judgments on the grammaticality of the linguistic facts. This single feature makes it distinct from the printed grammars. Hyper grammars are extremely useful from the point of learning, teaching and as reference material.

The design features are borrowed from the hypertext format but conceived as a computationally cognitive model. The contents are being developed from both the published and unpublished sources carefully selected and rewritten in the hypertext format.

2. The Contents:

The content of Tamil Hyper grammar has two main components, viz. 1. the description of grammar in hypertext format and 2. the applicational aspect of the Tamil Language as a language manager.

2.1. The Tamil Grammar:

The grammar part includes a number of comprehensive descriptive notes on certain linguistic facts of the Tamil Language. It is conceived in terms of a Computational Grammar. It deals with the Orthography, the design features of Tamil script, the orthographic syllable, the information on the frequency distribution of written syllables etc.

As part of the Tamil morphology, we have information on Tamil categories viz. nouns, adjectives, verbs, adverbs, numerals, pronouns etc. In each of these, there is information regarding the setting up of paradigm types and a list of paradigmatic forms under each category. One can access information regarding the most frequent hundred words, five thousand words and ten thousand words in terms of their frequencies, and communicative contribution to the coverage in Tamil Texts. As regards to the frequency of Tamil characters and syllables as they occur in the 3 million-word corpus, one can find the relevant information. One of the most important and crucial is the lexical component. A number of bilingual dictionaries like Tamil-Hindi, Tamil-Kannada, Tamil-Tamil, Tamil-Oriya, Tamil-Marathi, Tamil-English and English-Tamil - are included. Originally these dictionaries are conceived as bilingual and bi-directional dictionaries initially created using the most frequently occurring words ensuring the coverage.

2.2. The Tamil language manager:

This is the most crucial component of Tamil Hyper grammar. It involves the actual functions of the practical aspect of the grammar outlined above. As said earlier, the grammatical description is only a statement about the competence of a native speaker - about his/her language. In order to make it to simulate the grammar, it should involve a working generator, analyzer, parser and lexical accessor, etc. Currently the Tamil language manager includes a word form generator, a morphological analyzer and lexical accessor among others.

a. The Morphological Analyzer:

The word analyzer incorporated here is intended to analyze the Tamil words in terms of the lexical root/stem, its category, the paradigm type and the inflectional or derivational affixes attached to it. A morphological analyzer (Morph) engine essentially learns from a morphological lexical database of a particular language. The functional coverage and efficacy of the engine is greatly dependent on the structure and the organization of the database. The database of Tamil Morphological Analyzer comprises of inflectional database and the root dictionary. These data comprise purely linguistic information of the language, which are processed subsequently to enable using it in morphological analysis. It uses the Word and Paradigm Model of analysis.

The Organization of the Linguistic data for Morph:

(i) The paradigmatic-data

The term Paradigm refers to an exhaustive set of morpho-syntactically related word forms of a given lexeme. Based on the inflection, six distinct morphological categories are identified and the paradigms are created. They include the major and minor categories of words.

(a) The major word classes which are productive and open class categories (new members are added from time to time) can inflect with distinct but characteristic suffixes which explicit morpho-syntactic functions. The major word categories are listed as below,

1. Nouns
2. Verbs
3. Adjectives

(b) The distinct minor categories which are productive but considered as closed class categories (no new members are added) are listed below,

4. Pronouns
5. Numerals
6. Locative Nouns

The other class of words which are not fallen under the above categories are a list of idiosyncratic word forms. They cannot inflect for any functional categories. They come under functional categories of language with defective morphology. The following words are usually known as indeclinable and have no morphology to process.

1. Postpositions
2. Adverbs
3. Conjunctions
4. Interjections
5. Particles

The above words are listed as 'Avy' (avyayas) in the dictionary.

(ii) Root Dictionary

Root Dictionary is a vast collection of lexemes which contains words, their categoral information and their suitable paradigms. It includes a certain number of minimally distinct words in the semantic system of a language. This is typically called as lexicon without semantics.

Input : a valid word form

Output :

1. Root
2. Lexical Category
3. Paradigm type
4. Morphological Category

(The output may be one or more analysis)

Input and Output Specifications in Tamil:

Input:

- 1 koyampuwwUr¹
- 2 iraNtAvawu
- 3 mikappeVriya
- 4 wamilYYaka
- 5 mAnakarakam
- 6 Akum

Output:

- | | | | |
|---|--------------|-----|--|
| 1 | koyampuwwUr | unk | <fs af='koyampuwwUr,n,n,sg,3,d,0,0'> |
| 2 | iraNtAvawu | unk | <fs af='iraNtu,num,,sg,,o,Avawu,Avawu'> |
| 3 | mikappeVriya | unk | <fs af='mikappeVriya,adj,any,any,any,,0,0'> |
| 4 | wamilYYaka | unk | <fs af='wamilYYakam,n,n,sg,3,o,0,0'> |
| 5 | mAnakarakam | unk | <fs af='mAnakaram,n,n,sg,3,d,0,0'> |
| 6 | Akum | unk | <fs af='Aku,v,n,sg,3,,pp,pp'>
<fs af='Aku,v,n,pl,3,,pp,pp'> |

b. Word form Generator:

A Tamil word-form synthesizer enables a user to generate Tamil word forms. The user is prompted to select some choices leading to the generation of the desired word. This is extremely useful to the learners of Tamil as second language. Such uses can interactively generate the requested word in Tamil.

The Morphological Generator of Tamil is based on Word and Paradigm Method. It is built using the feature values, suffix informations with add or delete rules and the root word dictionary with its category and paradigm. It uses the Machine Learning techniques to generate the word form from the given input. The basic resources required for present word synthesizer:

Feature Value:

It contains the category, its possible morpho-syntactic properties. It has five values, each viz., category, gender, number, person and the affix. For instance,

“v m sg 3 nw”

The above is an example for the verb for generating third person singular masculine past tense verb as such *vanwAnY* 'he came'.

Suffix information and synthesis rule set:

This is generated from the paradigms and its feature values. It contains the rules for words based on their morpho-phonemic processes. It has four columns delimited by comma. For instance, to generate

'marafkalYE'
maram + kalY +E
Eng: tree + plural+ Accusative case
the suffix information table consists,
“Eylakf,m,maram,89”

Whereas the first is an inversed suffix of 'fkalYE' which is to be added, the second is the word which has to be deleted from root and the third is the name of the paradigm as such the word behaves in its morphophonemic process and finally the row number of the feature value file.

Lexicon:

Lexicon consists of the root words of Tamil, its category and the name of the paradigm based on its phonological behaviour in its inflection. For instance,

'aNi,v,varE' Eng: put on, verb, draw
 Here *aNi*, the verb act morpho-phonemically as *varE*.
aNi-nw-AnY as in *varE-nw-AnY* 'root-PST-3p.sg.m'
aNi-kirY-AIY as in *varE-kirY-AIY* 'root-PRS-3p.sg.f'
aNi-v-ArkaLY as in *varE-v-ArkaLY* 'root-FUT-3p.sg.m'

- Input :**
1. Root
 2. Lexical Category
 3. Morphological Category

Output : a valid word form

Input and Output Specifications in Tamil:

Input:

- 1 kampar NNP <fs af='kampar,n,m,sg,3,0,0,0'>
- 2 irAmAyaNam NN <fs af='irAmayaNam,n,n,sg,3,o,E,E'>
- 3 iyarYrYu VM <fs af='iyarYrYu,v,m,sg,3,,nw,nw'>

Output:

- 1 **kampar** NNP <fs af='kampar,n,m,sg,3,0,0,0'>
- 2 **irAmAyaNawwE** NN <fs af='irAmayaNam,n,n,sg,3,o,E,E'>
- 3 **iyarYrYinYAr** VM <fs af='iyarYrYu,v,m,sg,3,,nw,nw'>

c. Dictionary :

The Tamil-Telugu bilingual dictionary is built based on the concepts available in the language. It differs from the conventional dictionary which lists words but not concepts. Here, the lexeme(s) are related to each other on the basis of the concept i.e. the idea of ontological entity. The dictionary which is based on concepts is a better one for obtaining a concise and effective lexicon which can be used in many NLP applications.

d. Machine Translation System :

The development of Machine Translation is one of the most challenging tasks of the Natural Language Processing Applications. The development of Machine Translation (MT) System which translates texts from Telugu to Tamil and vice-versa (Bi-directional) are incorporated here. This MT system was developed as part of IL-ILMT consortium project funded by the Government of India at CALTS, University of Hyderabad. This Machine Translation system uses Transfer Based Approach. The System's Architecture is divided into three stages i.e. Source language Analysis module (SLA), Source language to Target language Transfer module (SL-TL) and Target language generation module (TLG).

(i) Telugu-Tamil Machine Translation system:

The *crucial* tools used in Telugu-Tamil Machine Translation system includes,

a. *Source Language Analysis*

- Telugu Sandhi Splitter
- Telugu Morphological Analyzer
- Telugu POS Tagger
- Telugu Chunker
- Telugu NER (Named Entity Recognizer)
- Telugu Parser

b. *Source Language- Target Language Analysis*

- Telugu-Tamil Transfer Grammar Module
- Telugu-Tamil Multi Word Expression Module
- Telugu-Tamil Lexical Transfer Module

c. *Target Language Analysis*

- Tamil Agreement Module
- Tamil Word form Generator

(ii) Tamil-Telugu Machine Translation:

The *crucial* tools used in Tamil-Telugu Machine Translation system includes,

a. *Source Language Analysis*

- 1) Tamil Sandhi Splitter
- 2) Tamil Morphological Analyzer
- 3) Tamil POS Tagger
- 4) Tamil Chunker
- 5) Tamil NER (Named Entity Recognizer)
- 6) Tamil Parser

b. *Source Language- Target Language Analysis*

- 1) Tamil-Telugu Transfer Grammar Module
- 2) Tamil-Telugu Multi Word Expression Module
- 3) Tamil-Telugu Lexical Transfer Module

c. *Target Language Analysis*

- 4) Telugu Agreement Module
- 5) Telugu Word form Generator

3. Conclusion:

The Tamil Hyper Grammar thus described here is the most significant development in the recent applications of Natural language processing of the Tamil language to be used as teaching, learning as well as reference grammar for all kinds of language users.

¹ Transliteration Scheme using wx-notation:

Tamil Orthography :

a A i I u U eV e E oV o O H

k f c F t N w n p m y r l v IY Y IY rY nY j s h R

Telugu Orthography :

a A i I u U q Q e V e E o V o O M H

k K g G f c C j J F t T d D N w W x X n p P b B m y r r Y l I Y I Y Y v S R s h

References :

- [1] Arden A.H. 1891. A Progressive Grammar of the Tamil Language. Chennai: The Christian Literature Society.
- [2] ILMT Consortium. 2007. ILMT SRS and Functional Specifications (mimeo). Hyderabad.
- [3] Parameswari K. 2009. An Improvized Morphological Analyzer for Tamil: A case of Implementing an open source platform Apertium. Unpublished M.Phil. Thesis. Hyderabad: University of Hyderabad.
- [4] Ramaswamy, Vaishnavi. 2003. A morphological Analyzer for Tamil. Unpublished Ph.D. Thesis. Hyderabad: University of Hyderabad.
- [5] Uma Maheshwar Rao, G. 2002. A Computational Grammar of Telugu. (Mimeo) Hyderabad: University of Hyderabad.
- [6] Uma Maheshwar Rao, G. 2005. Telugu Hyper Grammar. (Mimeo and Electronic form) Hyderabad: University of Hyderabad.
- [7] Uma Maheswar Rao G, Amba P. Kulkarni and Christopher M. 2007. Functional Specifications of Morphology (mimeo). Hyderabad.
- [8] Uma Maheswar Rao G. and Christopher M. 2010. Word Synthesizer Engine. In Morphological Analyzer and Generators. Mona Parakh (ed.) Page 73-81. Mysore; CIIL.
- [9] Uma Maheswar Rao G. and Parameswari K. 2010. On the Description of Morphological Data for Morphological Analysers and Generators: A case of Telugu, Tamil and Kannada. In Morphological Analyzer and Generators. Mona Parakh (ed.) Page 114-123. Mysore; CIIL.

A Tamil - Telugu Bi-directional Machine Translation System

Christopher M | efthachris@gmail.com

Krupanandam N | nityakrupa@yahoo.co.in

Parameshwari K | parameshkrishnaa@gmail.com

Uma Maheshwar Rao G | guraohyd@gmail.com

Vijaya Bharathi D | vijaya.anthony@gmail.com

Centre for Applied Linguistics & Translation Studies

University of Hyderabad, Hyderabad-500046.

Abstract

We present the development of Machine Translation (MT) System which translates texts from Telugu to Tamil and vice-versa (Bi-directional). This MT system was developed as part of IL-ILMT consortium project funded by Govt. of India at CALTS, University of Hyderabad. This Machine Translation system uses Transfer Based Approach. System's Architecture is divided into three stages i.e. Source language Analysis module (SL), Source language to Target language Transfer module (SL-TL) and Target language generation module (TL). The computational Modules that are used in the building of this system were developed mainly by CALTS-UoH, IIIT-H and AUKBC research teams. We also use the statistical open source engine i.e. CRF++ for POS-Tagging, Chunking and Named Entity Recognizer (NER). Hence the architecture is a hybrid one.

1. Introduction: The development of Machine Translation (MT) is one of the most challenging tasks of Natural Language Processing Applications. In MT there are a number of methods that are being practiced all over the world, chiefly, they are *Direct Methods*, *Interlingual Methods*, *Transfer Based Approach* and a combination of these beside the statistical and corpus based methods. Tamil and Telugu are two closely related languages, which belong to the same i.e. the Dravidian language family. Even though they belong to the same language family, still they exhibit a considerable amount of diversity at every level viz. morphological, syntactic, semantic and lexical levels. Keeping these in mind, building a Machine Translation System for this language pair using Transfer based Method can be non-trivial and challenging. The present paper discusses the successful implementation of the Transfer Based Approach to the Machine Translation (MT) System for the Telugu-Tamil pair. This bi-directional Telugu-Tamil MT system is one of the nine pairs of Indian Language to Indian Language Machine Translation Systems (IL-ILMT) planned to be developed by the Consortium of IL-ILMT constituted by the DIT, MIT, Govt. of India. The system is an assembly of various linguistic modules run on specific engines whose output is sequentially maneuvered and modified by a series of modules till the output is generated. The most crucial linguistic modules include, a Morphological Analyzer (MA), Parts of Speech Tagger (POS-T), a

Simple Parser (SP), the Transfer Grammar Component (TG), a Lexical Transfer module consisting of a Bilingual Dictionary and a Conceptual Dictionary, an Agreement module (AGR) and a Morphological Generator (Wordgen). The system is already built and is now being tested and evaluated. The presentation would involve the demonstration of a randomly selected text from the internet.

2. Module Level details of the System:

2.1. The Format:

The entire system works on a unique standard format called the Shakti Standard Format (SSF). The multi-column format vividly represents input and output of in each module throughout the system. This is especially designed to represent the different kinds of linguistic analyses, as well as different levels of analysis. The two kinds of analyses are : 1. Constituent level analysis and 2. Relational-Structure level analysis. The former is used to store simple phrase level analysis and the latter for storing relations between the simple parses. Feature structures are used to store attribute-value pairs for a phrasal node as well as for a word or a token. Attribute value pairs store relations in different columns. The following is a description of the column format in SSF:

Column 1 stores the node address, mainly for human readability. Column 2 stores the word or word-group input. The symbol “(” represents the start of the word or word-group and the symbol “)” to represent the end of the word or word-group. Column 3 stores the chunk name or the POS tag of the words occurred in the sentence. Column 4 stores the Morphological information (feature structures) of the words. Column 5, 6, and 7 store the gender, number and person feature values respectively. Column 8 stores the oblique or direct nature of the stem in case of nouns and 9 the case marker in case of nouns and tense in case of verbs. Column 10 store the exact suffix representing the features represented in 5-9.

2.2. Source Analysis:

Tokenizer: The tokenizer converts a text into a sequence of tokens (words, punctuation marks, etc.) within the Shakti Standard Format.

a. Morphological analyzer (MA): A Morphological Analyzer analyzes and identifies the root and the grammatical features of the word. Word and Paradigm based approaches have given good success rates for Indian languages (Rao et. al. 2007). The computational module used in this MT system provides about 20-30 features values for each word, in which 8 are mandatory viz. root, lexical category, gender, number, person, case, case marker or tam and suffix. The lexical categories are divided into 9, they are noun (n) , verb (v), adjective (adj), pronoun (pn), adverb (adv), postposition (psp), number (num), nouns of space and time (NST) and indeclinable (Avy) (Rao et.al. 2007).

1	himAlayAlu	unk	<fs af='himAlayaM,n,,pl,,d,0,0'> <fs af='himAlayaM,n,,pl,,d,vu,vu'>
2	sahaja	unk	<fs af='sahajaM,n,,sg,,o,ti,ti'> <fs af='sahajudu,n,,sg,,o,ti,ti'>
3	sixXaMgA	unk	<fs af='sixXaM,n,,sg,,,gA,gA'>
4	erpaddAyi	unk	<fs af='erpadu,v,n,pl,3,,A,A'>
5	.	unk	<fs af='.,punc,,,,,'>

b. Parts of speech tagger (POS-T): Part of speech tagging is the process of assigning a unique part of speech to each word (token) in the sentence. This process helps in identifying the role of each word

(token) in a sentence. There are number of approaches, such as rule-based, statistics based, transformation-based etc. which use for POS tagging. Here we propose to use statistical techniques on a Gold standard manually developed tagged text (follows ILMT Tagset, ILMT 2007).

c. Chunker: Chunking involves identifying non-recursive combinations of word groups involving nouns (NP), verbs (VGF/VGNF), adjectives (JJP) and adverbs (RBP) etc. in a given sentence. Here we use statistical methods to identify and chunk tags in a sentence (following ILMT Tagset, ILMT 2007).

```

1      ((      NP
1.1  himAlayAlu NN    <fs af='himAlayaM,n,,pl,,d,0,0'> | <fs af='himAlayaM,n,,pl,,d,vu,vu'>
      ))
2      ((      NP
2.1  sahaja  NN    <fs af='sahajaM,n,,sg,,o,ti,ti'> | <fs af='sahajudu,n,,sg,,o,ti,ti'>
      ))
3      ((      RBP
3.1  sixXaMgA  RB    <fs af='sixXaM,n,,sg,,,gA,gA'>
      ))
4      ((      VGF
4.1  erpaddAyi  VM    <fs af='erpadu,v,n,pl,3,,A,A'>
4.2  .          SYM  <fs af='.,punc,,,,,'>
      ))

```

d. Named Entity Recognizer (NER): The identification, recognition and tagging of proper nouns such as names of persons and organizations (ILMT 2007) is achieved by this module.

```

1 (( NP <fs af='himAlayaM,n,,pl,,d,0,0' head='himAlayAlu' ENAMEX TYPE='LOCATION'
SUBTYPE_1='LANDSCAPES'>
1.1  himAlayAlu  NN    <fs af='himAlayaM,n,,pl,,d,0,0' name='himAlayAlu'>
      ))
2 (( NP <fs af='sahajaM,n,,sg,,o,ti,ti' head='sahaja' vpos='vib1'>
2.1  sahaja  NN    <fs af='sahajaM,n,,sg,,o,ti,ti' name='sahaja'>
      ))
3 (( RBP <fs af='sixXaM,n,,sg,,,gA,gA' poslcat='NM' head='sixXaMgA'>
3.1  sixXaMgA  RB    <fs af='sixXaM,n,,sg,,,gA,gA' poslcat='NM' name='sixXaMgA'>
      ))
4 (( VGF <fs af='erpadu,v,n,pl,3,,A,A' head='erpaddAyi'>
4.1  erpaddAyi  VM    <fs af='erpadu,v,n,pl,3,,A,A' name='erpaddAyi'>
4.2  .          SYM  <fs af='.,punc,,,,,'>
      ))

```

e. Simple parser (SP): Identifies and names Thematic relations between a verb and its participant noun in the sentence, based on the Computational Paninian Grammar framework (Bharathi et.al 1995).

```

1      ((      NP      <fs af='himAlayaM,n,,pl,,d,0,0' head="himAlayAlu" drel=k1:4 name=1 ENAMEX
TYPE="LOCATION" SUBTYPE_1="LANDSCAPES">
1.1    himAlayAlu  NN      <fs af='himAlayaM,n,,pl,,d,0,0' name="himAlayAlu">
      ))
2      ((      NP      <fs af='sahajaM,n,,sg,,o,ti,ti' head="sahaja" vpos="vib1" name=2>
2.1    sahaja     NN      <fs af='sahajaM,n,,sg,,o,ti,ti' name="sahaja">
      ))
3      ((      RBP     <fs af='sixXaM,n,,sg,,,gA,gA' head="sixXaMgA" name=3 poslcat="NM">
3.1    sixXaMgA   RB      <fs af='sixXaM,n,,sg,,,gA,gA' poslcat="NM" name="sixXaMgA">
      ))
4      ((      VGF     <fs af='erpadu,v,n,pl,3,,A,A' head="erpaddAyi" name=4>
4.1    erpaddAyi  VM      <fs af='erpadu,v,n,pl,3,,A,A' name="erpaddAyi">
4.2    .          SYM    <fs af=',punc,,,,,'>
      ))

```

f. Transfer Grammar (TG): Wherever, the source language does not have an equivalent structure in the target language, a structural transformation is required to convert the source language structure into an acceptable target language structure. Such cases can be found at all levels wherever divergence occurs between the source and the target language. TG Module contains rules which convert the parsed structure of the source language into the desired structure in the target language giving the acceptable target structures.

```

1      ((      NP      <fs af='boVrra,n,,sg,,,0_0_A' head="boVrrA" drel=k1:9 name=3>
1.1    boVrrA     NNP     <fs af='^@boVrrA,n,,sg,,,0_0_A' ENAMEX/TYPE="LOCATION" name="boVrrA"
SUBTYPE_1="PLACE">
      ))
2      ((      NP      <fs af='guha,n,,pl,,d,0,0' head="guhalu" drel=k2:9 name=4>
2.1    kuke       NN      <fs af='^@kuke,n,,pl,,d,0,0' name="guhalu">
      ))
3      ((      NP      <fs af='10,num,,,,,' head="10" name=5>
3.1    10         QC      <fs af='10,num,,,,,' name="10" NUMEX TYPE="COUNT">
      ))
4      ((      NP      <fs af='^@ANtu,n,,pl,,o,ti,ti' vpos="vib2" head="eYla" name=6>
4.1    lakRala    QC      <fs af='lakRa,n,,pl,,o,ti,ti' poslcat="NM" NUMEX/TYPE="COUNT">
4.2    ANtu       NN      <fs af='^@ANtu,n,,pl,,o,ti,ti' name="eYla">
      ))
5      ((      NP      <fs      af='kriwaMnAtivi,n,n,sg,3,,0,0'      head="kriwaMnAtivi"      name=7
poslcat="NM">

```

5.1 kriwaMnAtivi NN <fs af='kriwaMnAtivi,n,n,sg,3,,0,0' poslcat="NM" name="kriwaMnAtivi">
 5.2 . SYM <fs af='.,punc,,,,,'>
))

g. Multi-Word Expression Transfer (MWE): Multi-Word Expression module involves identification and transfer of frequently used non-compositional phrases, compounds, reduplicatives, etc. from the source language to the target language.

1 ((NP <fs af='himAlayaM,n,,pl,,d,0,0' head="himAlayAlu" drel=k1:4 name=1>
 1.1 himAlayAlu NN <fs af='himAlayaM,n,,pl,,d,0,0' name="himAlayAlu">
))
 2 ((NP <fs af='sahajaM,n,,sg,,o,ti,ti' head="sahaja" vpos="vib1" name=2>
 2.1 0 NN <fs af='^@0,n,,sg,,o,ti,ti' name="sahaja">
))
 3 ((RBP <fs af='sixXaM,n,,sg,,,gA,gA' head="sixXaMgA" name=3 poslcat="NM">
 3.1 iyarYkERB <fs af='^@iyarYkE,n,,sg,,,gA,gA' name="sixXaMgA" poslcat="NM">
))
 4 ((VGF <fs af='erpaddu,v,n,pl,3,,A,A' head="erpaddAyi" name=4>
 4.1 erpaddAyi VM <fs af='erpaddu,v,n,pl,3,,A,A' name="erpaddAyi">
 4.2 . SYM <fs af='.,punc,,,,,'>
))

h. Lexical transfer (LT): Root words identified by the morphological analyzer are looked up in a bilingual dictionary for the target language equivalent using concept substitution including function words.

1 ((NP <fs af='இமாலயம்,n,,pl,,d,0,0' head='himAlayAlu' drel='k1:4' name='1'>
 1.1 இமாலயம் NN <fs af='இமாலயம்,n,,pl,,d,0,0' name='himAlayAlu'>
))
 2 ((NP <fs af='இயற்கை,n,,sg,,o,ti,ti' vpos='vib1' head='sahaja' name='2'>
 2.1 0 NN <fs af='0,n,,sg,,o,ti,ti' name='sahaja'>
))
 3 ((RBP <fs af='தயார்,n,,sg,,,ஆக,gA' head='sixXaMgA' name='3' poslcat='NM'>
 3.1 iyarYkERB <fs af='iyarYkE,n,,sg,,,ஆக,gA' name='sixXaMgA' poslcat='NM'>
))
 4 ((VGF <fs af='ஏற்படு,v,n,pl,3,,ந்த்,A' head='erpaddAyi' name='4'>
 4.1 ஏற்படு VM <fs af='ஏற்படு,v,n,pl,3,,ந்த்,A' name='erpaddAyi'>
 4.2 . SYM <fs af='.,punc,,,,,'>
))

i. Agreement (Agr): Performs checking and reconstructing gender-number-person agreement between the subject and the predicate in the target sentence, ensuring proper agreement.

j. Vibhakti Splitter or Complex Inflection Splitter (VBS): Separates complex cases of inflections involving postpositions and auxiliary verbs ensuring proper word generation.

```

1      ((      NP      <fs af='AnwirA,n,n,sg,,d,0,0' head=AMXra drel=k1:5 name=1>
1.1    0      NNP     <fs      af='0,n,n,sg,,d,0,0'      ENAMEX/TYPE=LOCATION      name=AMXra
SUBTYPE_1=PLACE>
      ))
2      ((      NP      <fs af='pirawecam,n,n,sg,,d,0,0' head=praxeS name=2>
2.1    Anwirapirawecam      NN      <fs af='Anwirapirawecam,n,n,sg,,d,0,0' name=praxeS>
      ))
3      ((      NP      <fs      af='hEwarApAw,n,n,sg,,,E,ni'      vpos=vib1      head=hExarAbAxni      drel=k2:5
name=3>
3.1    hEwarApAw      NN      <fs      af='hEwarApAw,n,n,sg,,,E,ni'      ENAMEX/TYPE=LOCATION
name=hExarAbAxni SUBTYPE_1=PLACE>
      ))
4      ((      RBP     <fs      af='walEnakaram,n,,sg,,,Aka,gA'      head='rAjaXAnigA'      name='4'
poslcat='NM'>
4.1    walEnakaram      RB      <fs af='walEnakaram,n,,sg,,,Aka,gA' name=rAjaXAnigA poslcat=NM>
      ))
5      ((      VGF     <fs af='peVrYu,v,fn,sg,3,,i_iru+nw,i' vpos='tam1_2' head='kaligi' name='5'>
5.1    peVrYu      VM      <fs af='peVrYu,v,any,any,any,,i,i' name=kaligi>
5.2    iru+nw      VAUX   <fs af='iru,v,fn,sg,3,,nw,'>
5.3    .          SYM     <fs af='.,punc,,,,,'>
      ))

```

k. Word generator (WG): This module takes root words and their associated grammatical features, selects appropriate suffixes and concatenates them into well formed word forms.

```

1      ((      NP      <fs af='ஆந்திரா,n,n,sg,,d,0,0' head='ஆம்தர்' drel='k1:5' name='1'>
1.1      NNP     <fs      af='0,n,n,sg,,d,0,0'      ENAMEX/TYPE='LOCATION'      name='ஆம்தர்'
SUBTYPE_1='PLACE'>
      ))
2      ((      NP      <fs af='பிரதேசம்,n,n,sg,,d,0,0' head='ப்ரதேஸ்' name='2'>
2.1    ஆந்திரபிரதேசம்NN      <fs af='ஆந்திரபிரதேசம்,n,n,sg,,d,0,0' name='ப்ரதேஸ்'>
      ))
3      ((      NP      <fs af='ஹைதராபாத்,n,n,sg,,,ஐ,ni' head='ஹைதராபாத்நி' vpos='vib1' drel='k2:5'
name='3'>

```

```

3.1 ஹைதராபாதை NN <fs af='ஹைதராபாத்,n,n,sg,,,ஐ,ni' ENAMEX/TYPE='LOCATION'
name='ஹைதராபாத்' SUBTYPE_1='PLACE'>
))
4 (( RBP <fs af='தலைநகரம்,n,,sg,,,ஆக,GA' head='ராஜதாநிகா' name='4' poscat='NM'>
4.1 தலைநகரமாக RB <fs af='தலைநகரம்,n,,sg,,,ஆக,GA' name='ராஜதாநிகா' poscat='NM'>
))
5 (( VGF <fs af='பெறு,v,fn,sg,3,,இ_இரு+ந்த்,i' head='கலிகி' vpos='tam1_2' name='5'>
5.1 பெற்று VM <fs af='பெறு,v,any,any,any,,இ,i' name='கலிகி'>
5.2 இருந்தது VAUX <fs af='இரு,v,fn,sg,3,,ந்த்,'>
5.3 . SYM <fs af='.,punc,,,,,'>
))

```

l. Post Processing: Enables if any unacceptable sequences of words to be modified in terms of more acceptable structures of the target language.

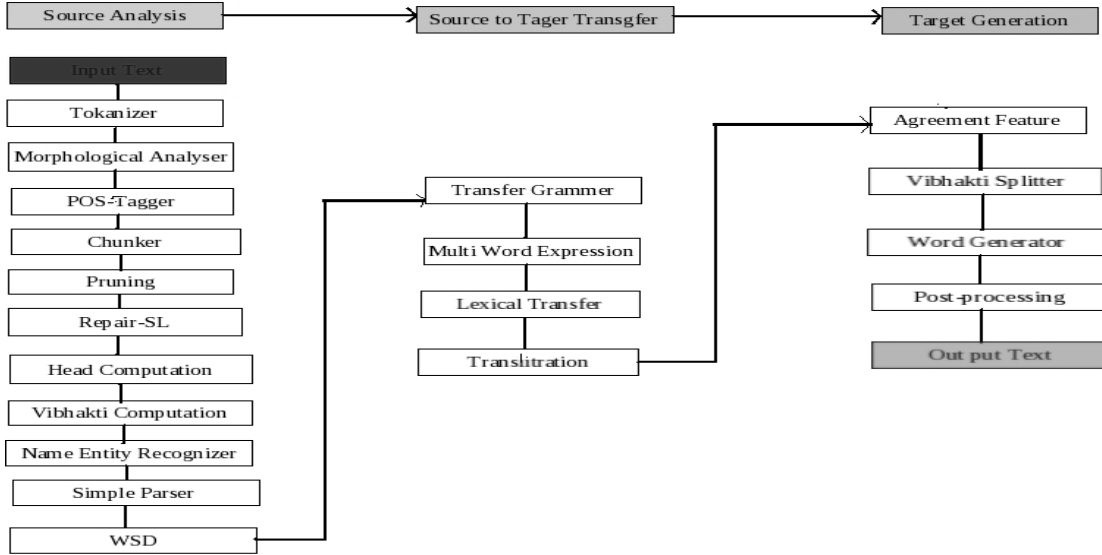
```

1 (( NP <fs af='ஆந்திரா,n,n,sg,,d,0,0' head='ஆம்தர்' drel='k1:5' name='1'>
1.1 NNP <fs af='0,n,n,sg,,d,0,0' ENAMEX/TYPE='LOCATION' name='ஆம்தர்'
SUBTYPE_1='PLACE'>
))
2 (( NP <fs af='பிரதேசம்,n,n,sg,,d,0,0' head='ப்ரதேஸ்' name='2'>
2.1 ஆந்திரபிரதேசம்NN <fs af='ஆந்திரபிரதேசம்,n,n,sg,,d,0,0' name='ப்ரதேஸ்'>
))
3 (( NP <fs af='ஹைதராபாத்,n,n,sg,,,ஐ,ni' head='ஹைதராபாத்' vpos='vib1' drel='k2:5'
name='3'>
3.1 ஹைதராபாதை NN <fs af='ஹைதராபாத்,n,n,sg,,,ஐ,ni' ENAMEX/TYPE='LOCATION'
name='ஹைதராபாத்' SUBTYPE_1='PLACE'>
))
4 (( RBP <fs af='தலைநகரம்,n,,sg,,,ஆக,GA' head='ராஜதாநிகா' name='4' poscat='NM'>
4.1 தலைநகரமாக RB <fs af='தலைநகரம்,n,,sg,,,ஆக,GA' name='ராஜதாநிகா' poscat='NM'>
))
5 (( VGF <fs af='பெறு,v,fn,sg,3,,இ_இரு+ந்த்,i' head='கலிகி' vpos='tam1_2' name='5'>
5.1 பெற்றிருந்தது VM <fs af='பெறு,v,any,any,any,,இ,i' name='கலிகி'>
5.3 . SYM <fs af='.,punc,,,,,'>
))

```

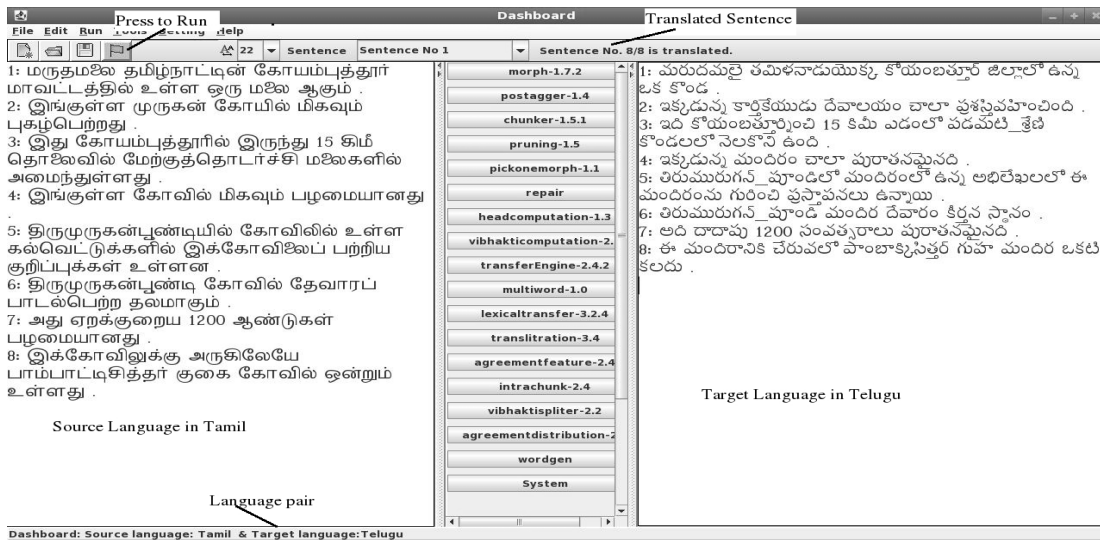

3. Conclusion:

The architecture of this system is based on analyze-transfer-generate paradigm. The flow of the input sentence in the system is given in fig:1.



All the modules have been integrated on the dashboard, a tool, where the data flow in the pipeline is configured. This ensures speed, since it uses shared memory. This MT system demonstrated here is a completely automated translation system without involving human interference for the first time involving Tamil. Though the current system is built for the tourism domain, it can be extended to any other domain. The system can be used to translate web pages or text material from books, magazines, newspapers etc. written in standard language. It runs on Linux platform with Apache-2.0 server. The browser used for the online translation can be Firefox 1.0.4, IE 6.0 or Mozilla 1.7.8.

Sample Input and Output of the System in Dashboard.



References:

1. Akshar Bharathi, Vineet Chaitanya and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. New Delhi:Prentice Hall of India.
2. Uma Maheswar Rao G. and Christopher M. 2010. *Word Synthesizer Engine*. In *Morphological Analyzer and Generators*. Mona Parakh (ed.) Page 73-81. Mysore; CIIL.
3. Uma Maheswar Rao G. and Parameshwari K. 2010. *On the Description of Morphological Data for Morphological Analysers and Generators: A case of Telugu, Tamil and Kannada*. In *Morphological Analyzer and Generators*. Mona Parakh (ed.) Page 114-123. Mysore; CIIL.
4. ILMT Consortium. 2007. *ILMT SRS and Functional Specifications (mimeo)*. Hyderabad.
5. Uma Maheswar Rao G, Amba P. Kulkarni and Christopher M. 2007. *Functional Specifications of Morphology (mimeo)*. Hyderabad.

Certain issues in the Development of Telugu - Tamil Machine Translation

A view from the lexicon

Parameswari K | parameshkrishnaa@gmail.com

Uma Maheshwar Rao G | guraohyd@gmail.com

Krupanandam N | nityakrupa@yahoo.co.in

Lavanya J | jlavanyacse08@gmail.com

Christopher M | efhachris@gmail.com

Center for Applied Linguistics and Translation Studies

University of Hyderabad, Hyderabad - 500046.

Abstract:

Machine Translation (MT) is one of the interesting and challenging tasks of Natural Language Processing. In any Machine Translation system, understanding the pair of languages involved are vital. The present work focuses on certain issues in the development of Telugu-Tamil Machine Translation from the point of the languages involved and the dictionaries that are used in Telugu-Tamil Machine Translation System which are unique since they are based on concept. The paper deals with the compilation of concept based dictionary for Machine Translation purpose and also deals with the divergences arise due to the differences in the lexemes of Telugu and Tamil.

1. Introduction:

Tamil, the South Dravidian Language and Telugu, the South Central Dravidian language are major languages of South India. The Machine Translation between Telugu-Tamil is a best example case taken for the development of MT since there is a great demand for the Translation of texts of each of these languages. Normally, Machine Translation is a challenging task where computers take over the task of translating one language into another. Though the languages involved viz. Telugu - Tamil are closely related, exhibit a number of dissimilarities in their linguistic behavior thus making the task a non trivial one.

The paper deals with the issues in the development of an automatic Telugu-Tamil Machine Translation System which is being developed under the project of IL-IL MT at CALTS, University of Hyderabad as part of the Consortium of Indian Languages to Indian languages Machine Translation Systems funded by DIT, Ministry of Information Technology, Government of India.

The lexical resources are essential for building any Machine Translation system. The Lexicon used in the building of Telugu-Tamil MT is one of the Machine Readable Dictionary types, which differs from printed conventional dictionaries of everyday use. The conventional dictionary is usually meant for defining and providing description about a lexeme. However, the concept based dictionary which is currently used contains lexemes without any encyclopedic knowledge.

2. Concept based Dictionary or Synset:

A concept is an idea which is language specific and based on the ontology of lexemes in languages. The concept Based dictionary is a component of a multilingual dictionary developed for 11 languages: English, Hindi, Bengali, Marathi, Punjabi, Urdu, Tamil, Kannada, Telugu, Malayalam and Oriya (Cf. Mohanty et.al.) by different Indian Institutions and are used in NLP applications. The greatest advantage of using synset is the conceptually related words are grouped under a single concept and the equivalents in the target language along with linkages provided. Here, Hindi is used as a pivot language and other synsets in other languages are built on the principle of translational equivalence. The Telugu-Tamil Machine Translation System uses Telugu and Tamil synsets which are developed by CALTS' NLP group and AU-KBC NLP group respectively. A lexeme used to express a concept in a language may not have the same meaning in all the contexts. The same lexeme may be found in different contexts expressing different meanings or concepts. For instance, a lexical item in Telugu corresponds to one or more lexical (sense's) items in Tamil.

In Telugu, the word *kuttu*¹ is translated in Tamil as,

- a. *kati* in the context of *cIma kuttu* 'to bite as an ant',
- b. *wE* in the context of *battalu kuttu* 'to stitch clothes' ,
- c. *kuwvu* in the context of *ceVvulu kuttu* 'to pierce ears'.

Here the question of providing an appropriate equivalent for '*kuttu*' requires word sense disambiguation. The concept which is the central point of this lexeme can help to avoid this problem. The dictionary which we are proposing as a suitable one for Machine Translation is of concept centered one rather than of one to one lexical matching. A word X in a language is taken as a concept, and the conceptually related words of X are provided as W1,W2,...Wn. The hierarchy of frequency is followed in an ascending order of giving equivalents. The links (L) are created between the source language and the target language lexemes. The concept Dictionary is used to perform a lexical transfer of the following:

(a) Situation (1) One to One :

Here a single lexical item is linked with a corresponding lexical item in Tamil.

Ex: X (sw1/L1 <--> tw1)

(where X is a context with category, sw is a source word, L is link, tw is target word)

Ex:

Telugu :

ID :: 7350

CAT :: NOUN

CONCEPT :: ఎవరోకైనా అప్పమ ఇచ్చోచోనప్పమడు లోదా బోయాంకు మూదలైన నాటోలో కూడబోటోన డబ్బుకో బదులుగా ఆ సమయం వరకు ఇచ్చోచో నోశోచోత ధనము

EXAMPLE :: "శోయాం వడోడోకో డబ్బులు ఇనోతాడు"

SYNSET-TELUGU :: వడోడో /TAM1

.....>

This is the link to

Tamil.

Tamil:	
ID	:: 7350
CAT	:: NOUN
CONCEPT	:: வட்டி
EXAMPLE	:: "வங்கியில் வாங்கிய கடன் தொகைக்கான வட்டி குறைந்துள்ளது."
SYNSEM-TAMIL	:: வட்டி

(b) Situation (2) Many to One :

Here multiple lexemes are displayed with linkages with a single lexical item

Ex: X(sw1/L1, sw2/L1, sw3/L1, sw4/L1, sw5/L1 <--> tw1)

Telugu :	
ID	:: 73
CAT	:: NOUN
CONCEPT	:: అంతర్గతంగా కలిగి ఉన్న క్రోయ
EXAMPLE	:: "అందరినీ అందంగా ఉండే భవం ఉన్నదా"
SYNSEM-TELUGU	:: భవం/TAM1, భవార్యం/TAM1, భవన/TAM1, అర్థం/TAM1, తృప్త్యం/TAM1
Tamil:	
ID	:: 73
CAT	:: NOUN
CONCEPT	:: இப்படிப்பட்டது அல்லது இப்படிப்பட்டவர் என்பதை அறிவதற்கான அம்சம்
EXAMPLE	:: "மனிதனிடம் மனித தன்மை காணப்படும்."
SYNSEM-TAMIL	:: தன்மை

(c) Situation (3) One to Many :

Here a single lexeme is linked with multiple lexical items.

Ex : X(sw1/L1 <--> tw1, tw2,tw3)

Telugu :	
ID	:: 12019
CAT	:: NOUN
CONCEPT	:: ఒక వస్తువును దగ్గరకు లాగే స్థితి
EXAMPLE	:: "అయినకొంతలకు ఆకర్షణ శక్తి ఉంటుంది/తన కళ్లలో"
SYNSEM-TELUGU	:: ఆకర్షణ/TAM1
Tamil:	
ID	:: 12019
CAT	:: NOUN
CONCEPT	::
EXAMPLE	::
SYNSEM-TAMIL	:: கவர்ச்சி, வசீகரம்,, ஈர்ப்பு

(d) Situation (4) Many to Many :

Here many lexical items are linked with many in the target side.

Ex: X(sw1/L1, sw2/L2, sw3/L3, sw4/L4, sw5/L5 ,sw6/L6, sw7/L7 <--> tw1,tw2,tw3,tw4,tw5,tw6)

Telugu:

ID :: 12833

CAT :: VERB

CONCEPT ::

EXAMPLE :: ""

SYNSET-TELUGU :: పరారంభించు/TAM1, మౌదలాపూటలు/TAM2,
లసద్యోయి/TAM3, ఆరంభించు/TAM4, పరారంభించుయి/TAM5, ఆరంభించు/TAM6,
లసద్యోయి/TAM7

Tamil :

ID :: 12833

CAT :: VERB

CONCEPT :: எழுப்பு, ஆரம்பி

EXAMPLE :: "நடு நடுவே அவன் ரமாவின் திருமணத்தைப் பற்றிய பேச்சை
எழுப்பினான்"

SYNSET-TAMIL :: ஆரம்பி, தொடங்கு, எழுப்பு, ஆரம்பி, ஆரம்பம்_செய், துவங்கு,
எழுப்பு

The dictionary uses the categories like Nouns, Verbs, Adjectives, Adverbs, Pronouns, Numerals, NST and Indeclinables such as Classifiers, Quantifiers, Interjections, Quotatives, Particles and Conjunctions. Other than this, a whole list of functional words like Case Markers and Tense, Aspectual and Model markers are also included in the dictionary. A bilingual dictionary, which is also a concept based one is used as a stand-by along with the synset dictionary in case of failing.

3. Divergences of Telugu and Tamil from the point of lexicon:

A translation divergence may occur when the underlying concept or "gist" of a sentence is distributed over different words for different languages. According to Dorr (1990), divergences are cross-linguistic distinctions in which the natural translation of one language into another results in a very different that of the original. She proposes seven types of divergent categories comprising of Thematic, Promotional, Demotional, Structural, Conflational, Categorical and Lexical Divergences. This classification of Dorr (1990) on Machine Translation Divergence is taken as a base and is tried to map it with the Telugu-Tamil Machine Translation System. The paper focuses on three divergences due to lexical aspects of the languages involved in translation.

3.1 Conflational Divergence :

It occurs when the sense conveyed by a single word is expressed by two or more words in one of the languages. For instance, Telugu uses 'snAnaM ceVyyi' for 'bath' whereas it is expressed by 'kulYi' in Tamil.

II.a. TEL: nenu snAnaM ceswAnu.
'I bathing do-FUT-1p.sg'
TAM: nAnY kulYippenY.
'I bath-FUT-1p.sg'
ENG: I will bathe.

The Conflational Divergence is mainly carried out by the Multi Word Expression Module. Here the collocative words are given equivalents in the respective language. Multi-word expressions are a set of collocations of words which are often come with a non compositional semantics which otherwise could not be resolved. These forms are sequences of two or more words generally express a co-occurrence meaning. Telugu-Tamil Multi-Word Expression Module is built up with the database which consummates the words of co-occurrence. The root form of the two or more sequences of words are used in the database. For instance, the following expression of Noun (N) and Verb (V) is carried out during the Telugu-Tamil Translation:

1. N N --> N N
uwwara praxeS, uwwirap pirawecam

Since 'uwwaraM' in isolation may mean either 'the north' or 'a letter', but in the context of the word indicating the name of a State, it needs to be listed.

2. N V --> N V
veru ceVyyi, pirivinYE ceVy

Here the word 'veru' may mean 'separation' and 'root'. But when it is followed by a verb like 'ceVyyi' it means 'to separate'.

3. N N N --> N N N
calana ciwra pariSrama, wirEp patac cafkam

Here the cinema is expressed in Telugu as 'calana ciwraM' i.e, 'motion picture' whereas in Tamil it is 'screen picture'.

4. N N --> 0 N
sahajaM sixXaM, 0 iyarYkE

The term 'natural product' is expressed by two words in Telugu whereas in Tamil it is one.

5. N V --> 0 V
vidixi ceVyyi, 0 wafku
xAdi ceVyyi, 0 wAkku

In Telugu, the intensifier compounds involve two words to express a single intensified form of the concept denoted by the nouns of temporal/spatial category whereas in Tamil by the corresponding reduplication of the head noun.

6. REDUP NST ---> REDUP NST
moVtta moVxata, muwanY muwal

As described above lexemes of special cases such as phrases, idioms which are multi word expressions are taken care by the MWE module before the processing enters into the lexical transfer module.

Input :

```

<Sentence id="1">
1 (( NP <fs af='samuxraM,n,,sg,,o,ti,ti' head="samuxra" vpos="vib1" name=1>
1.1 samuxra NN <fs af='samuxraM,n,,sg,,o,ti,ti' name="samuxra" ENAMEX
TYPE="LOCATION" SUBTYPE_1="LANDSCAPES">
))
2 (( NP <fs af='wIraM,n,,sg,,d,0,0' head="wIraM" drel=k7p:4 name=2>
2.1 wIraM NN <fs af='wIraM,n,,sg,,d,0,0' name="wIraM">
))
</Sentence>

```

Output :

```

<Sentence id="1">
1 (( NP <fs af='samuxraM,n,,sg,,o,ti,ti' head="samuxra" vpos="vib1" name=1>
1.1 0 NN <fs af='^@0,n,,sg,,o,ti,ti' ENAMEX/TYPE="LOCATION" name="samuxra"
SUBTYPE_1="LANDSCAPES">
))
2 (( NP <fs af='wIraM,n,,sg,,d,0,0' head="wIraM" drel=k7p:4 name=2>
2.1 katarYkarE NN <fs af='^@katarYkarE,n,,sg,,d,0,0' name="wIraM">
))
</Sentence>

```

3.2 Categorical Divergence :

Changes in category creates categorical divergence. It is due to the mismatch between the Parts of Speech Categories of the words involved in the pair of Languages. For instance, the word cAla in Telugu is ambiguously used as an adjective as well as an intensifier. However, in Tamil two distinct categories of words are used as shown below:

- III.a. TEL: cAlA puswakAlu unYnYAYi.
'a lot book-pl being-3.p.pl.n'
- TAM: nirYEya puwwakafkalY ulYIYanYa.
'a lot book-pl being-3.p.pl.n'
- ENG: A lot of books are there.
-
- III.b. TEL: cAlA eVwwugA uMxi.
'very high being'
- TAM: mika uyaramAka ulYIYawu.
'very high being'
- ENG: It is very high.

Handling Strategy :

(i) Repair Role :

$\{[cAIA<\$cat>][N1.gA]\} \Rightarrow \{[cAIA<INT>][N1.gA]\}$
 $\{[cAIA<\$cat>][N1]\} \Rightarrow \{[cAIA<ADJ>][N1]\}$

3.3 Lexical Divergence:

It arises when there is a lack of exact lexical equivalent but structure presents a translational equivalence between the language pair. Here, the literal translation of the source language word is substituted by a corresponding translational equivalent to resolve the problem.

For instance, IV.a. TEL: nAku Iwa vaccu.
'me-DAT swimming come'
TAM: eVnYakku nIccal weVriyum.
'me-DAT swimming know-fut-3p.sg.n'
ENG: I know to swim.

Handling Strategy :

Transfer Rule :

$\{[N1ku][N20][vaccu]\} \Rightarrow \{[N1ku][N20][teri<3p.sg.n>]\}$

IV.b. TEL: AmeVku kadupu vacciMxi.
'She-DAT bellyt come-PST-3p.sg.n'
TAM: avalYukku karppam erYpattawu.
She pregnancy form-PRS-3p.sg.f'
ENG: She became pregnant

In the above example, the idiomatic sense of 'kadupu' is 'pregnant' which means 'the stomach'. But Tamil uses the term 'karppam' to express the same.

4. Conclusion: The Telugu- Tamil Machine Translation system is built by using the concept based dictionaries discussed above. The concept based dictionaries ensure the resolution of much of the disambiguation presented by the words in the lexical substitution in translation. The system is tested continuously by the native speaker of Tamil in order to validate its performance in the translation. The five scale Evaluation method of IL-IL MT is adopted for this purpose. The current comprehension of the outputs fall between **85-90%**.

¹ Transliteration Scheme using wx-notation:

Tamil Orthography :

a A i I u U eV e E oV o O H
k f c F t N w n p m y r l v IYY IY rY nY j s h R

Telugu Orthography :

a A i I u U q Q eV e E oV o O M H
k K g G f c C j J f t T d D N w W x X n p P b B m y r rY l IY IYY v S R s h

References:

- [1] Arden A.H. 1891. *A Progressive Grammar of the Tamil Language*. Chennai: The CLS.
- [2] Bhuvanewari . G. 2009. *Telugu-Tamil Machine Transaltion*. Unpublished Ph.D. Thesis, University of Hyderabad.
- [3] Dorr, Bonnie. 1990b. *Solving Thematic Divergence in Machine Translation*. In the Proceedings of the 28th Anual Conference of the ACL,127-134, University of Pittsburg, Pittsburg, PA.
- [4] Dorr, Bonnie . 1993. *Machine Translation: A View from the Lexicon*. Cambridge, Mass: The MIT Press.
- [5] Krishnamurti, Bh and Gwynn, J.P.L. 1985. *A grammar of modern Telugu*. New Delhi: OUP.
- [6] Mohanty Rajat K., Bhattacharya P and et. al. *Synset Based Multilingual Dictionary : Insights, Applications and Challenges*. :_www.cse.iitb.ac.in / ~pb/papers/ gwc08- multilingual- dictionary.pdf
- [7] Sangal Rajeev, Uma Maheshwar Rao G, Nagamma Reddy K, 1999. proceedings of the National Seminar of 'information Revolution and Indian Languages' , *Society for Computer Applications in Indian Languages : Hyderabad*.
- [8] Sinha, R.M.K., Thakur, A. 2005. *Translation Divergence in English-Hindi MT EAMT*, Budapest, Hungary.
- [9] Uma Maheshwar Rao, G. 2002. *A Computational Grammar of Telugu*. (Momeo). Hyderabad: University of Hyderabad.

EILMT: A Pan-Indian Perspective in Machine Translation

Hemant Darbari,

Executive Director, C-DAC, Pune, darbari@cdac.in

Anuradha Lele,

Group Co-ordinator, C-DAC, Pune, lele@cdac.in

Aparupa Dasgupta,

Team Co-ordinator, C-DAC, Pune, aparupa@cdac.in

Priyanka Jain,

Project Leader, C-DAC, Pune, priyankaj@cdac.in

Sarwanan, Amrita University, sarwanster@gmail.com

Abstract

To cut-across the language barrier and to encourage the language pluralism of morphologically complex languages [Sproat 1991], especially South-Asian languages [Krishnamurti et al. 1986] in India, a consortium mode robust Machine Translation system (MTS) that is able to raise the accuracy of generation is developed jointly by C-DAC, Pune and DIT, GOI. In Natural Language Processing (NLP) and Natural Language Understanding (NLU), Machine Translation plays a vital role in today's India for any sort of e-language processing and understanding by machine. In each of the quarter of electronic era of a multi-lingual community machine translation, information retrieval or speech processing becomes obligatory. This paper proposes to describe a hybrid based machine translation system from English to Indian languages. This paper also proposes the TAG based memory managed Machine Translation System [Joshi et al. 1981] aligning with other rule based, example based and statistical based Machine Translation System for English-Hindi, English-Urdu, English-Oriya, English-Bangla, English-Marathi and English-Tamil. EILMT has especially been designed to translate in platform independent modules. This is a proposed hybrid based thin-client/thick-server design; where users (clients) of this system use a standard browser to access the translation services of the server. We call this as a Pan-Indian perspective on Machine Translation. In this paper, we will explain the challenges faced and solution drawn at the various levels of architecture, language and linguistic computation. While building the Machine Translation System, we have taken care of the speed and accuracy of syntactically and morphologically diversified languages at modular and phases of EILMT system.

1.0 Introduction to Machine Translation

In present paper, we explain the challenges encountered to cope with the speed and accuracy of syntactically and morphologically diversified languages tested and developed for Machine Translation system based on consortium mode for English to Indian Languages in collaboration with C-DAC, Pune and DIT, Govt. of India. In 1629 the idea of Machine Translation evolved, when Rene Descartes proposed Universal Language. In 1954, the Georgetown experiment (1954) involved fully-automatic translation of sixty Russian sentences into English. In late 1980s, machine translation inclined to statistical models and example based models evolved gradually. And the Machine Translation system like Systran used by AltaVista search engine, METEO used at the Canadian Meteorological Centre, Example-based machine

translation proposed by Makoto Nagao and several other Hybrid based Machine Translation system came into existence. During the year 1990-91, DIT (Department of Information Technology) of Government of India initiated the TDIL (Technology for Development of Indian languages) project to encourage the Indian language processing in the area of IT. The institutions namely, C-DAC, Pune (MANTRA); NCST (now C-DAC, Mumbai; MATRA); IIIT-Hyderabad (Anusaaraka, and SHAKTI) and IIT-Kanpur (Anglabharati) have taken the Machine Translation System from English to Hindi to greater height by developing applications using cutting edge technology.

2.0 Introduction to EILMT

To overcome the language barrier and to encourage the language pluralism of morphologically complex languages [Sproat 1991], especially South-Asian languages [Krishnamurti et al. 1986] in India, a consortium mode robust Machine Translation system (MTS) that is able to raise the accuracy of generation is developed jointly by C-DAC, Pune and DIT, Govt. of India. It is domain specific Machine Translation system from the domain of tourism. This project is developed by 10 consortium institutes: C-DAC, Mumbai, IIIT-Hyderabad, IISc-Bangalore, IIT-Bombay, Jadavpur University - Kolkata, Amrita University - Coimbatore, IIIT-Allahabad, Banasthali Vidyapeeth - Banasthali, Utkal University - Bhubaneswar and C-DAC, Pune being the consortium leader. EILMT is a hybrid based Machine Translation system with TAG formalism (Tree Adjoining Grammar based MT developed by C-DAC, Pune), SMT (statistical based MT developed by C-DAC, Mumbai), ANALGEN (Rule based MT by IIIT-Hyderabad) and EBMT (Example based system developed by IISc, Bangalore). To measure the performance of aforementioned translation engines and evaluate the language pair wise translation accuracy, we represent here the internal testing carried out by consortium. The translation output accuracy on English-Hindi EILMT system of each of these aforementioned translation engines are given below. Following table data is the average score of engine for each sentence structure type:

Sentence Structure type	AnalGen (%)	EBMT (%)	SMT (%)	TAG (%)
Copula	88.75	42.50	66.25	87.50
Simple	58.75	60.00	57.50	95.00
Appositional	71.50	47.50	67.50	91.25
Relative Clause	75.00	53.75	55.00	95.00
That-Clause	62.50	51.25	60.00	92.50
Wh-Clause	56.66	38.33	53.75	63.33
Co-ordinate	65.00	32.50	78.75	95.00
Conditional	48.50	56.25	63.75	77.50
PP Initial	60.00	43.75	57.50	93.75
Adverb Initial	80.00	53.33	53.33	95.00
Gerundial	36.00	31.60	75.00	83.33
Participle	81.25	25.00	81.25	91.25
Infinitive	50.00	46.25	60.00	93.75
Discourse Connector	70.00	55.00	75.00	75.00

Table 1: Engine wise Translation output accuracy for English -> Hindi pair

Similarly, language pair wise translation accuracy on TAG translation engine was evaluated, whose approximate translation accuracy is as follow: for English-Hindi pair the translation accuracy is approximately 85%; for English-Urdu is approximately 75%; for English-Oriya is approximately 80%; for English-Bangla is approximately 70%; for English-Marathi is approximately 65%; and for English-Tamil is approximately 70%.

3.0 Introduction to EILMT Architecture: *The Challenges*

EILMT is a web-based Machine Translation system solution with a hybrid approach across six language-pairs from English to Hindi, Urdu, Oriya, Bangla, Marathi and Tamil. Along with four different machine translation engines, the Named Entity Recognizer [NER] and Word Sense Disambiguation [WSD] modules are developed by IIT, Mumbai. EILMT system architecture is represented in the following diagram:

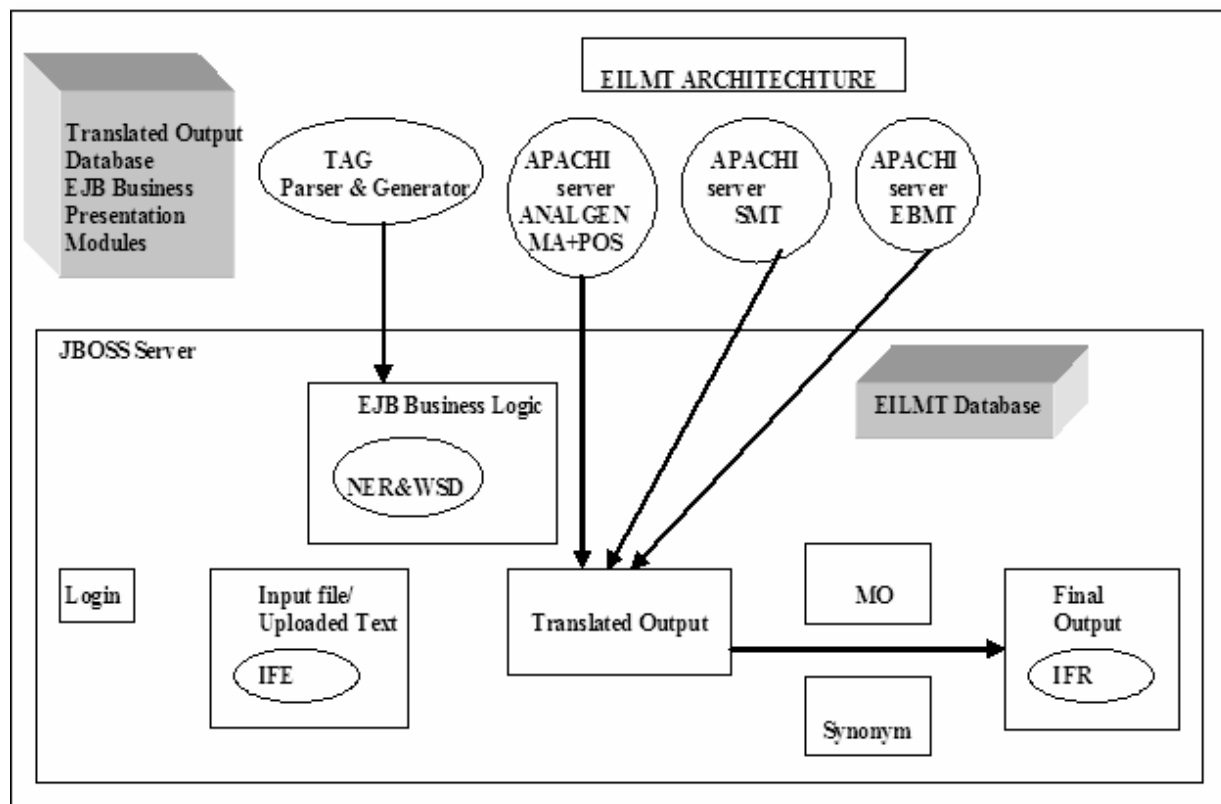


Diagram 1: EILMT system architecture

Basic system components of EILMT consortium are: User Log module; Pre-Processing module; four Translation Engines: AnalGen, EBMT, SMT & TAG for six language pairs; Post-Processing module; Collation and Ranking module; a compatible system with W3C; and Browser compatibility for IE, Mozilla, Firefox, Google Chrome, Apple Safari & Opera. (See *Annexure 1 B* for detailed EILMT system specifications).

3.1 Overall Architecture of EILMT

EILMT is a web based translation system accessed simultaneously with multiple users and requests. JBoss is the application server with robust database I/O file/exe and for rapid, transactional, secure and portable application EILMT is supported by EJB (Enterprise Java Bean). EILMT is designed on the line of centralized design where Internet clients submit their documents to a multi-core server where the parsing and generation is a spawning of multi-threaded embedding. Significantly, the outer layer thread connects to ANALGEN engine (implemented in PERL on Linux platform), another thread with SMT with server and the other with EBMT engine. And the Ranking module collates and rank the translation from the above mentioned translation engines.

EILMT system has been tested on multi core (8 core machine) machine for execution that has raised the system processing speed upto three times. Initially NER which is used in SMT system developed by C-DAC, Mumbai followed a Maximum Entropy Based Approach. This system had an accuracy of 81.33% on ConLL-2003 dataset. (Precision: 83.85%, Recall: 78.95% F-Measure: 81.33%). The current system uses two stages: SVMs followed by MEMMs. Using 2 phases, improved the accuracy to 93% (Precision: 92.56% Recall: 93.48% F-Measure: 93%).

3.2 Implementation of TAG Formalism

Tree Adjoining Grammar [Kroch and Joshi, 1985] is implemented for all 6 language-pairs in EILMT on TAG translation engine. The JAVA based TAG parser translates English documents to Hindi, Urdu, Oriya, Bangla, Marathi and Tamil. The significant feature of this parser is incremental parser that identifies the

- (a) clause or phrase on the basis of probable declarative clause boundary and,
- (b) after identifying clause boundary the TAG tree derivation structure identifies probable parent derivation to the nearest child derivation structure to give the final integrated derivational tree to the TAG Generator.

The TAG engine is enriched in such a way that it can process the parsing and generation for interrogative sentences, negation, gerundial construction, relative clause construction, and past & progressive participle etc. The pre-processing is controlled by supervised modules such as – syntactic TAG tree disambiguator module with optimized code and database-design written in regular expressions. Consider the following description of the incremental parser that has given modularity, extensionality and speed in the translation process of TAG engine. Probability of adjoining the parent derivations to a nearest probable child derivation is given by the following equation:

$Y = \{c(X)\}$; Where, X = Number of Child derivations, Y = Number of Parser Derivations, c = Combination,

Consider the sentence “The 18th century Bharatpur-Bird-Sanctuary, which is also known as the Keoladeo-Ghana-National Park, is famous as the most important bird breeding and feeding habitat of the world.”

Following is the parse derivation of clause (one of the clause):

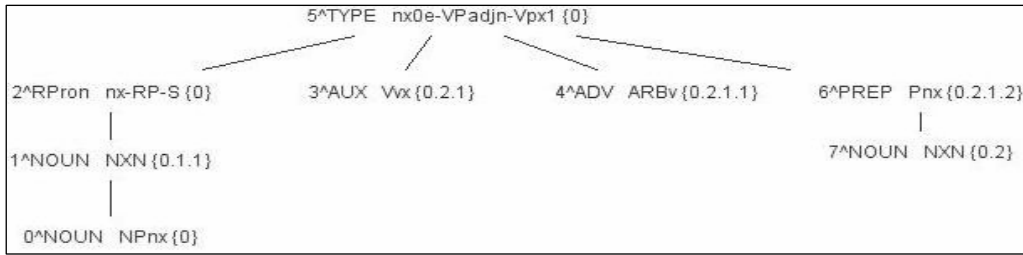


Diagram 2: Parser Derivation of clause - 1

Following is the complete Generated derivation (or derived Tree):

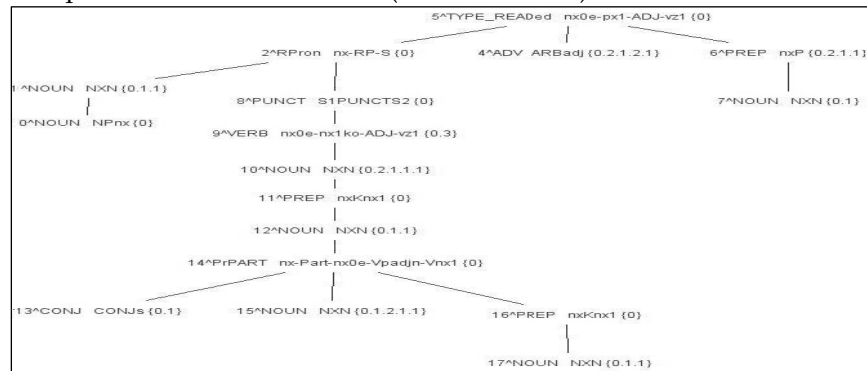


Diagram 3: Complete Generated derivation

4.0 Linguistic Diaspora of EILMT: Morphologically Diversified Languages

The English corpus of 15,200 sentences from tourism domain were collected, organized, vetted and aligned [Sinclair, J. 1991 and 2004] for all 6 language-pairs. India being a Linguistic Area [see Krishnamurti et al, 1986] in South-Asian sub-continent, both Indo-Aryan (eastern and western Indo-Aryan) and Dravidian language families with rich morphological heritage have its separate distinct linguistic identity at source-target TAG grammar, transfer grammar (a source-target link grammar), rule-normalizer, rules for morphological analysis and synthesis and transliteration and typing-tool rule. The stylistic trend observed in EILMT tourism corpus is: simple sentence (14.94% frequency of occurrence), copulative construction (3.49% frequency of occurrence), co-ordinate sentences (20.10% frequency of occurrence), appositional sentences (11.33% frequency of occurrence), various declarative clause structures (22% frequency of occurrence), gerundial constructions (.35% frequency of occurrence), conditional sentences (1% frequency of occurrence), discourse connector (0.77% frequency of occurrence) and infinitival sentences (9.03% frequency of occurrence). Thus, the parallel corpus created for all 6 language-pairs and features such as intelligibility, comprehensibility and fluency in translation are maintained to set a reference to the machine output, as E.M. Enquest has said very correctly that “Proper words in proper places creates styles”.

In Natural Language Processing (NLP) and Natural Language Understanding (NLU) [Terry Patten, 1985], Machine Translation plays a vital role in Indian sub-context for e-language processing by machine. In Eng-Hindi EILMT system, the localization of linguistic peculiarities of Hindi such as oblique formation, ergativity, marked-gender system, case-marking, direct-oblique pluralization etc. are handled in a controlled environment through morph-synthesizer, finite and non-finite generators, POS conversion rule

etc. Similarly, for other Indo-Aryan language pairs i.e., Eng-Urdu, Eng-Oriya, Eng-Bangla and Eng-Marathi the linguistic features such as, Perso-Arabic and Indic pluralization system, lexico-semantic peculiarities, copula drop, dropping of existential subject, post-position synthesis, synthesis of case-marking, emphatic clitic formation, usage of classifier, verb root alteration, strong and three level gender system, gender based noun synthesis, and compounding etc., [Bhattacharya, T et. al 1996; Krishnamurti, Bh. et al 1986; Selkirk 1982; Williams 1981] are incorporated through feature-based lexicon, ordered rule-based normalizer etc. Approximately 37,000 bilingual lexicon, 2000 phrasal lexicon, 97 TAG tree disambiguation rule, 125-150 source TAG trees, 215-230 target TAG trees, 800 transfer grammar mapping and 70-75 morph-synthesis rule are developed for each Indo-Aryan language-pairs. Following section will explain the linguistic challenges faced and language computing solutions drawn in EILMT system for syntactically and morphologically diversified and complex languages to raise the speed and translation accuracy:

4.1 Raising Translation Speed and Accuracy: Intermediate Solutions

a) Rectifying wrong POS tagging of Stanford tagger (version 1.6) through rule based POS tagging (See *Computational Linguistics*, volume 19, number 2, pp313-330.). Consider the following examples that states the internal POS conversion rule that rectifies the erroneous tagging output of Stanford tagger,

“Visit the Sheesh-Mahal or the Hall of Victory glittering with mirrors and ascend the Fort on elephant's back”

Stanford Output: [Visit@@@NN, the@@@DT, Sheesh@@@NNP, Mahal@@@NNP, or@@@CC, the@@@DT, Hall@@@NNP, of@@@IN, Victory@@@NNP, glittering@@@VGB, with@@@IN, mirrors@@@VBZ, and@@@CC, ascend@@@VB, the@@@DT, Fort@@@NNP, on@@@IN, elephant@@@NN, zxt@zxt@zxt@zxt@@@NN, back@@@RB]

Internal Pos Category String: [Visit@@VERB the-Sheesh-Mahal@@NOUN or@@CONJ] the-Hall@@ NOUN of@@PREP Victory@@NOUN glittering@@PrPART with@@PREP mirrors@@NOUN and@@ CONJ back@@ADV ascend@@TYPE_APPOINT the-Fort@@NOUN on@@PREP elephant@@NOUN zxt@zxt@zxt@zxt@@AS]

Apart from POS tagging, emotion and sense tagging is necessary in Machine Translation to capture the semantic anomaly of the natural language.

b) Chunking is an important part of shallow parsing level. It minimizes the number of tokens to be sent to the core parser, thus reducing the number of possible adjunctions and effected the translation time as well as the translation quality. We perform noun phrase chunking and verb group collation. Consider the following example of Chunking at level-1stage,

[The-Prince/NNP of/IN Wales/NNP Museum]/NNP ,/, [the-Jahangir-art-Gallery]/NNP ,/, [the-various-churches]/NNS ,/, temples/NNS and/CC shrines/NNS including/VBG [the/DT one/CD]/NNP of/IN [Haji-Ali]/NNP out/IN on/IN [an-island]/NN linked/VBN by/IN [a-causeway]/NN ,/, are/VBP worth/JJ [a-glimpse]/NN

Chunking at the level-2 stage:

[The-Prince-of-Wales-Museum]/NNP ,/, [the-Jahangir-Art-Gallery]/NNP ,/, [the-various- churches]/NNS ,/, temples/NNS and/CC shrines/NNS including/VBG [the-one]/NNP of/IN [Haji-Ali]/NNP

out/IN on/IN [an-island]/NN linked/VBN by/IN [a-causeway]/NN ,/, are/VBP worth/JJ [a-glimpse]/NN

c) We use a TAG (Tree Adjoining Grammar) [Joshi *et al.*, 1975] parser, and for that we have created a number of trees to represent structure of source and target languages. In this formalism each token is tagged with a POS tag/category, on the basis of which a set of possible tree tags are assigned to the token. This process is called tree tagging. A sentence as a string of tree-tagged tokens, are then sent to the parser. When a token in a sentence is tagged with a number of trees the parser is liable to produce multiple derivations, most of them being inappropriate. This reduces accuracy and speed. To eliminate this spurious derivation, or at least minimize them, we adopted the technique of TAG tree pruning. Our pruning module further disambiguates according to the syntactic context and helps in selection of TAG tree in more precise way. Accuracy and speed of the system, thus, was substantially improved.

d) To handle the synthesis of constructions in Indian Languages, morphological complexities of nouns and verbs and their inter-relationships, and the kaaraka formalism [Gangopadhyay, M. 1990] in a defined context plays a major role in noun or verb synthesis. Henceforth various categories at adjoining positions as an adpositional words like adjectives, post-positions (parasargas), and various particles like the avyayas etc. are also within this defined context. Basically post-positions, avyayas have modified-modifier [Aronoff, M. 1976] function adding more linguistic information to the end-users of the target language. The feature embedded morphological rules (and also sometimes gender agreement) written for the synthesizer can be seen through the synthesized output. Verbs in the language demand the karaka identities and the nouns fulfil the demands according to the yogyataa. And, in a defined context, nouns demand parasargas or post-position on a semantic account. Following diagram explains the synthesis process in EILMT system:

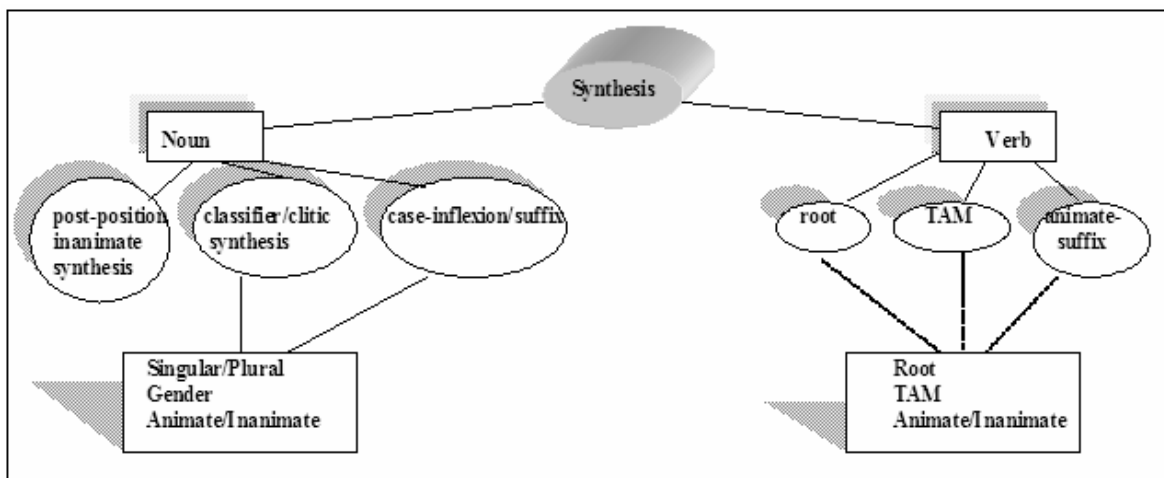
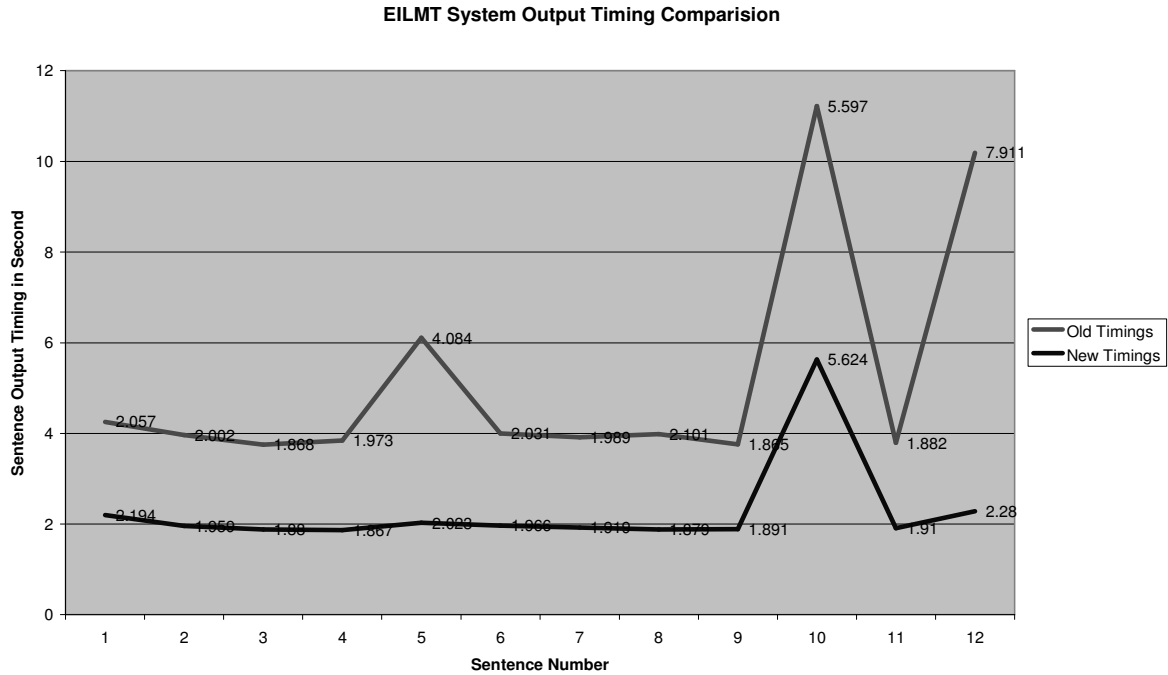


Diagram 4: Morph-synthesis process of EILMT system

Above mentioned points from 5 a) to d), the linguistic variations and complexities that are handled through pre-processing or post-processing generative modules have escalated translation accuracy and speed in a considerable way. Following graph represents the comparison of translation speed between old and new version of EILMT system (i.e., speed of translation before and after pruning and context disambiguation of the POS tagsets, TAG tree tagging and noun-verb synthesis). The above development at parsing and generation stages has raised the speed of translation in the latter (or new) version of EILMT:

Diagram 5: Comparison of speed of old and new version



5.0 English-Tamil EILMT System: an Overview

In English-Tamil EILMT system, special attention to Tamil morphological system has been given. As Tamil roots to Dravidian language family, being agglutinative language, the synthesis of finite and non-finite forms, synthesis of noun or noun group and gender based system has been catered through feature based lexicon, and noun and verb morph-synthesizer. In modern Tamil three types of words – noun, verb and itaiccol or particles are found. The noun indicates animate and inanimate categories (tiNai, is classified into uyartiNai and akRiNai). There are three genders in Tamil - masculine and feminine and neuter where masculine and feminine indicates singular number and neuter gender indicates plural number.

There are three persons in Tamil (first, second and third person). Case inflexion is prominent with suffixes in Tamil. Tamil being agglutinative in nature [See Varadarajan, Mu. 1988] is found to be different to parse and generate than the manner in which Indo-Aryan languages are generating in EILMT system. Approximately 35,000 bilingual lexicon, 92 phrasal lexicon, 97 TAG tree disambiguation rule, 125 source TAG trees, 127 target TAG trees, 147 transfer grammar mapping and 100 morph-synthesis rule are developed for English-Tamil version. Consider the following example from tourism domain with EILMT TAG output in Tamil:

English: Mother Earth' is kind in return. / **Tamil:** அன்னை பூமி திரும்புதலில் வகையாக இருக்கிறது

Following diagram represents the English-Tamil User Interface (with Tamil output):

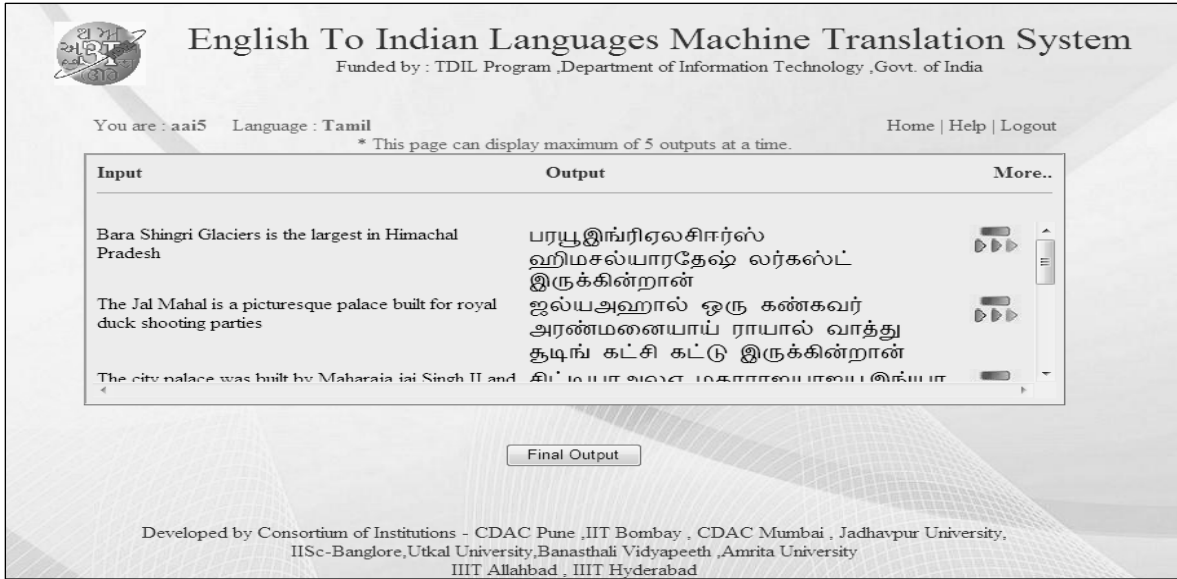


Diagram 6: English-Tamil User Interface output

5.1 Translation Accuracy of English-Tamil EILMT System

To evaluate the translation accuracy of English-Tamil system the score was evaluated through Subjective/Human Evaluation. The parameters for testing the translation accuracy of EILMT system for subjective/human evaluation are: POS tagging, P-Syntax, G-Syntax, Morph-Synthesis, Lexicon availability and phrase marking. We represent here the internal testing carried out by consortium on the test-report provided by the Testing Agency on EILMT alpha version 5.1 in the following bar-chart. [See Appendix I for English-Tamil output].

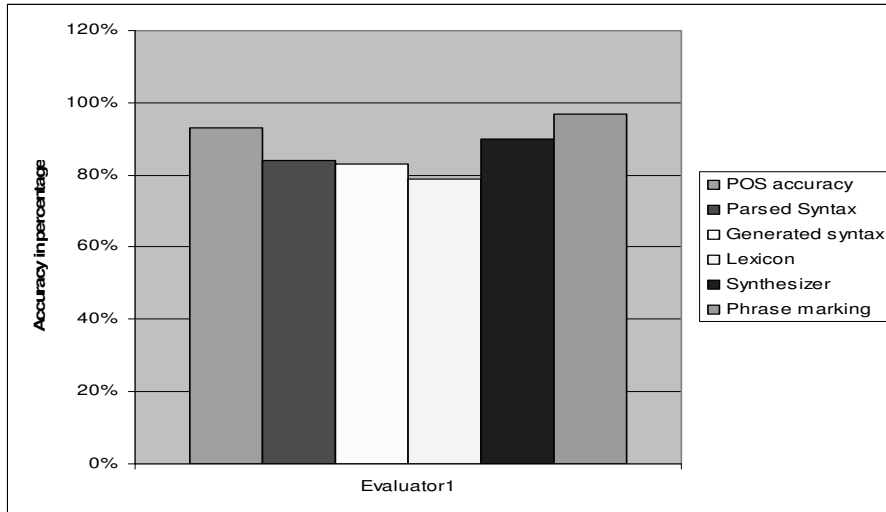


Diagram 7: Bar-Chart of Eng-Tamil translation output Evaluation

5.2 Scope of improvement for Eng-Tamil TAG EILMT system

Future improvement on English-Tamil EILMT system on the basis of development done for alpha version 5.1 is as follows:

- a) Re-framing and enhancing the Noun Collation module on the basis of Phrase Tagging and agglutinating character of Tamil.
- b) The process of new Tree set creation for the following structures: interrogative, imperative, negative sentences, handling of objects other than adverbial-synthesis.
- c) Enhancement of feature based linguistic rule-set in synthesis of synonym, the verb generator module for all the tenses and Bilingual lexicon correction.

6.0 Conclusion

All these above findings, research and implementations to EILMT system give a more productive and evolutionary ground. And this ground will definitely raise some critical questioning not only on machine translation but also text mining, data pruning, information extraction and retrieval, speech technology and in IL to IL information exchange and access. Thus the research and study on EILMT for Indian languages should be guided and formalized as following:

- a) Standardization of Indian tagset and considering the factor of morphologically rich language families and formal tagging, sense tagging and emotion tagging of the e-corpora available in Indian languages.
- b) Memory based parsing management to organize the multiple language with multiple domain. Further, memory managed MT will increase the system efficiency 15-20% more.
- c) And, feature-and-morphology based modules for morphologically rich Indian languages so that scope of this analysis and synthesis can be extended for reverse translation as-well.

7.0 Reference

1. Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge: MA: MIT Press
2. Bhattacharya, T. and P. Dasgupta 1996. *Classifiers, word order and definiteness in Bangla*. In V.S. Lakshmi and A. Mukherjee, ed. *Word order in Indian languages*. 73-94. Hyderabad: Booklinks.
3. Gangopadhyay, Malaya. 1990. *The Noun Phrase in Bengali: Assignment of Role and the*
4. *Kaaraka Theory*. Delhi: Motilal Banarsidass.
5. Joshi, Arvind, Bonnie Weber and Ivan Sag. 1981, *Elements of Discourse Understanding*. Cambridge University Press, New York.
6. Krishnamurti, Bh., C.P. Masica and A. K. Sinha (eds). (1986). *South Asian Languages: Structure,*
7. *Convergence and Diglossia*. Delhi: Motilal Benarsidass.
8. Kroch, T. and A. Joshi (1985). *The Linguistic Relevance of Tree Adjoining Grammar*. University of Pennsylvania. Department of Computer and Information.
9. Patten, Terry 1985. *A problem solving approach to generating text from systematic grammars*. Proceedings of 2nd Conference on European chapter of Association for Computational Linguistics. Geneva. Switzerland.

10. Sinclair, J. 1991. *Corpus, concordance, collocation*. Tuscan Word Centre, Oxford: Oxford University Press
11. _____. 2004. *Developing Linguistic Corpora: A Guide to good practice*. Oxford: Oxford University Press
12. Selkirk (1982). *The Syntax of Word*. MIT Press.
13. Vardarajan, Mu. 1988. *A History of Tamil Literature*. Translated from Tamil by E. Sa. Viswanathan 1-17. Sahitya Academy. New Delhi.
14. AAI Group, C-DAC, Pune. 2009. *EILMT Progress Report*. Submitted to DIT. Govt. of India. New Delhi

ANNEXURE I

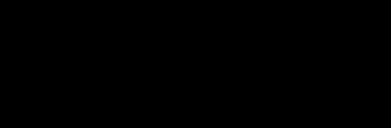
(Evaluation of EILMT system Version 5.1 Translation output for English-Tamil language pair)

English Sentence	Translated output (E-T)	Analysis of Translation	Type of Structure
The Ganga Golden Jubilee Museum has a large collection of pottery, paintings, carpets, coins, and armory.	கங்கா கோல்டன் ஜூபிலி மியூசியத்துக்கு மண்பாண்டத்தொழில், ஓவியங்கள், கம்பளங்கள், நாணயங்கள் மற்றும் படைக்கலனின் ஒரு பெரிய சேகரிப்பு இருக்கின்றது	POS= 100% P syntax = 100% G-syntax= 100% Lexicon= 100% Phrase Marking= 100% Synthesizer= 100%	Simple + Copula (Possessive form)
The shops are full with colorful items which include handicraft items, precious stones, textiles, Minakari items, jewellery, Rajasthani paintings, etc.	கடைகள் கைவினை வகைகள், மதிப்புமிக்க கற்கள், ஜவுளிகள், மினாக்கரி வகைகள், நகைக் கடை, ராஜஸ்தானி ஓவியங்கள், முதலியவையை உள்ளடக்குகிற வண்ணமயமான வகைகளுடன் முழுமையானவாக இருக்கின்றன	POS= 100% P syntax = 100% G-syntax= 100% Lexicon= 100% Phrase Marking= 100% Synthesizer= 100%	Relative Clause (subordinate clause)
Balsamand Lake & Palace, an artificial lake, is a splendid spot and was built in 1159 AD.	பல்சமந் ஏரி & அரண்மனை, ஒரு செயற்கையான ஏரி, ஒரு அழகான இடமாக இருக்கிறது மற்றும் 1159 ஆDஇல் கட்டப்பட்டது	POS= 100% P syntax = 100% G-syntax= 100% Lexicon= 95% Phrase Marking= 100% Synthesizer= 100%	Appositional(complement + initial)
The picturesque Kangra valley has several spots that offer mahaseer river carp.	கண்கவர் காங்ரா பள்ளத்தாக்குக்கு மஹாசீர் ஆறு குறைகூறை அளிக்கிற பல இடங்கள் இருக்கின்றன	POS= 100% P syntax = 100% G-syntax= 100% Lexicon= 80% Phrase Marking= 100% Synthesizer= 100%	(Co-ordinate) + That Complement

Visitors found the villagers dancing to the tune of folk music.	பார்வையாளர்கள் ிஃபோக் இசையின் ராகத்துக்கு டன்சின்க் கிராமத்தவர்களை கண்டுணர்கிறன	POS= 100% P syntax = 100% G-syntax= 100% Lexicon= 80% Phrase Marking= 100% Synthesizer= 100%	Gerundial
Udaipur is known for its beautiful lakes, well structured palaces, lush green gardens and temples but the major attractions of this place are the Lake Palace and the City Palace.	உதம்ப்பூர் அதனுடைய அழகான ஏரிகளுக்காக அறியப்படப்படுகின்றது , அரண்மனைகளை நன்றாக கட்டிடஅமைப்பை உருவாக்குக்கிறது , இந்த இடத்தின் வளமான அடர்ந்த பச்சை தோட்டங்கள் மற்றும் கோயில்கள் ஆனால் முக்கியமான கவர்ச்சிகள் லேக் பேலஸ் மற்றும் சிட்டி பேலஸாக இருக்கிறன	POS= 80% P syntax = 70% G-syntax= 70% Lexicon= 70% Phrase Marking= 100% Synthesizer= 100%	Discourse connector
Unless you are accustomed to horse riding, a daylong camel ride will be tiring.	இல்லாவிடில் நீங்கள் குதிரை சவாரிக்கு பயிற்றுக்கப்பட்டது , ஒரு நாள்முழுவது ஒட்டகம் சவாரியை சேர்வுறுத்துத்துக் கொண்டிருக்கும்	POS= 80% P syntax = 70% G-syntax= 70% Lexicon= 70% Phrase Marking= 100% Synthesizer= 70%	Conditional
Today, Mumbai is the country's financial and cultural centre, it is also home to a thriving film industry.	இன்று , மும்பை நாட்டின் பொருளாதார மற்றும் நாகரிகமான மையமாக இருக்கிறது , அது ஒரு ஆக்கவளமுறு மென்படலம் தொழிற்சாலைக்கு இருப்பிடமாக உம் இருக்கிறது	POS= 100% P syntax = 100% G-syntax= 100% Lexicon= 80% Phrase Marking= 100% Synthesizer= 90%	Complex sentence with Relative clause (Hidden) complement
Soon enough, prince Jahangir was born to his Hindu wife Jodha Bai.	சீக்கிரமாக போதுமானதாக , இளவரசன் ஜஹான்கிர் அவனுடைய இந்து மனைவி ஜோடா பாய்க்கு பிறப்பெடுக்கப்பட்டான்	POS= 100% P syntax = 90% G-syntax= 90% Lexicon= 80% Phrase Marking= 100% Synthesizer= 100%	Adverbial clause initial

10

கணினியில் தமிழ்த் தட்டச்சு



தமிழுக்கும் ஆங்கிலத்திற்கும் தரப்படுத்திய புதிய விசைப்பலகை அமைப்பு

ஆ.ரா.அமைதி ஆனந்தம் பி.ஈ, எம்.ஐ.ஈ,
Email: amaithianantha@yahoo.co.in

ஆயுள் உறுப்பினர், கணித்தமிழ்ச் சங்கம்,
ஆயுள் உறுப்பினர், இந்திய சாலைக் கழகம்,
தலைமை வரைதொழில் அலுவலர் (நெ) ஓய்வு, நெடுஞ்சாலைத்துறை, தமிழ்நாடு அரசு,
மூத்த குடிமகன், 2/7, 11 வது தெரு, சாந்தி நகர், ஆதம்பாக்கம்.சென்னை – 600088.

முன்பெல்லாம், தட்டச்சு இயந்திரத்தில், தட்டச்சு செய்வதற்கு விரல் அழுத்தம் (finger pressure) தேவைப்பட்டதால், விரல்கள் அழுத்தம் கொடுக்கும் (finger pressure) திறனுக்கு ஏற்றவாறு எழுத்துக்களின் இருப்பிடம் அமைக்கப்பட்டு தரப்படுத்தி உலகமெங்கும் பின்பற்றப்படுகின்றது. தற்போது தட்டச்சு இயந்திரத்திற்குப் பதிலாக கணினி வந்துவிட்டது. கணினியில் தட்டச்சு செய்வதற்கு விரல் அழுத்தம் (finger pressure) தேவைப்படவில்லை. விரல் தொடுதல் (finger touch) மட்டும் போதும். அதனால் விரல்கள் அழுத்தம் கொடுக்கும் (finger pressure) திறனுக்கு ஏற்றவாறு எழுத்துக்களின் இருப்பிடம் அமைக்க வேண்டிய அவசியம் இப்போது இல்லை. மேலும் அவ்வாறு அமைத்திருப்பது தேவை யில்லாமல் பயிற்சி கொடுக்க வேண்டியுள்ளது. எனவே தமிழுக்கு அகர வரிசைப்படி எழுத்துக்களை அமைத்து புதிய விசைப்பலகை அமைப்புகள் உருவாக்கப்பட்டு தரப்படுத்த வேண்டிய அவசியம் தற்போது உணரப்பட்டு வருகின்றது. அதேபோல், ஆங்கிலத்திற்கும் அகர வரிசைப்படி எழுத்துக்களை அமைத்து புதிய விசைப்பலகை அமைப்புகள் உருவாக்கப்பட்டு தரப்படுத்த வேண்டிய அவசியம் தற்போது உலகெங்கும் உணரப்பட்டு வருகின்றது.

எனவே, ஏற்கனவே உள்ள ஆங்கில விசைப்பலகை அமைப்போடு. திருத்திய தமிழ் 99 விசைப்பலகை அமைப்பு, புதிய தமிழ் விசைப்பலகை அமைப்பு இரண்டு வகைகள் மற்றும் புதிய ஆங்கில விசைப்பலகை அமைப்பு ஒரு வகை ஆக மொத்தம் 5 விசைப்பலகை அமைப்புகள் கீழ்க்காணுமாறு தரப்பட்டுள்ளன.

1. புதிய விசைப்பலகை அமைப்பு (ABCDEF விசைப்பலகை – ஆங்கிலம்)

ABCDEFGHIJ	10	மேல் வரிசை சாதா
abcdefghIj	10	மேல் வரிசை தூக்கல்
KLMNOPQRS	9	நடுவரிசை (தூக்கல்)
Klmnopqrs	9	நடுவரிசை (சாதா)
TUVWXYZ	7	கீழ் வரிசை (தூக்கல்)
Tuvwxyz	7	கீழ் வரிசை (சாதா)
மொத்தம்	52	எழுத்துக்கள்

2. தற்போதுள்ள விசைப்பலகை அமைப்பு (QWERT விசைப்பலகை - ஆங்கிலம்)

QWERTYUIOP	10	மேல் வரிசை சாதா
Qwertyulop	10	மேல் வரிசை தூக்கல்
ASDFGHJKL	9	நடுவரிசை (தூக்கல்)
asdfghjkl	9	நடுவரிசை (சாதா)
ZXCVBNM	7	கீழ் வரிசை (தூக்கல்)
zxcvbnm	7	கீழ் வரிசை (சாதா)
மொத்தம்	52	எழுத்துக்கள்

1. புதிய விசைப்பலகை அமைப்பு - 1 (அஆஇஈஉஊ விசைப்பலகை - தமிழ்)

ஸ (இடது மூலை)	(வலது மூலை) ஷ	2	மேற்பக்கம் (சாதா)
அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ		12	மேல் வரிசை (சாதா)
ஹ		12	நடுவரிசை (தூக்கல்)
க ங ச ஞ ட ண த ந ப ம ய			நடுவரிசை (சாதா)
ஃ		9	கீழ் வரிசை (தூக்கல்)
ர ல வ ழ ள ற ன ஐ			கீழ் வரிசை (சாதா)
(குறிப்பு 2 தூக்கல் விசை மட்டும்) மொத்தம்		35	முதன்மை எழுத்துக்கள்

2. புதிய விசைப்பலகை அமைப்பு - 3 (கசடதபற விசைப்பலகை - தமிழ்)

ஸ (இடது மூலை)	(வலது மூலை) ஷ	2	மேற்பக்கம் சாதா
ஒ ஏ ஊ ஈ ஆ	ங ஞ ண ந ம ன ஐ	12	மேல் வரிசை சாதா
	ஹ	12	நடுவரிசை (தூக்கல்)
ஒ எ உ இ அ	க ச ட த ப ற		நடுவரிசை (சாதா)
	ஃ	9	கீழ் வரிசை (தூக்கல்)
ஒள ஐ	ய ர ல வ ழ ள		கீழ் வரிசை (சாதா)
(குறிப்பு 2 தூக்கல் விசை மட்டும்) மொத்தம்		35	முதன்மை எழுத்துக்கள்

3. திருத்திய விசைப்பலகை அமைப்பு (தமிழ் 99 விசைப்பலகை - தமிழ்)

ஸ (இடது மூலை)	(வலது மூலை) ஷ	2	மேற்பக்கம் சாதா
ஒ ஒ ஏ ஈ ஆ	ள ற ன ட ண ச ஞ	12	மேல் வரிசை (சாதா)
	ஹ	1	நடுவரிசை (தூக்கல்)
ஒ எ உ இ அ	க ப ம த ந ய	11	நடுவரிசை (சாதா)
	ஃ	1	கீழ் வரிசை (தூக்கல்)
ஒள ஐ ழ வ ங ல ர ஐ		8	கீழ் வரிசை (சாதா)
(குறிப்பு 2 தூக்கல் விசை மட்டும்) மொத்தம்		35	முதன்மை எழுத்துக்கள்

அகரம் சேர்ந்த உயிர்மெய் எழுத்துக்களுக்கு “அ” விசை தேவைப்படவில்லை. அதனால் அந்த “அ” விசையை அணைத்து மெய் எழுத்துக்களுக்கு புள்ளி இருவதற்குப் பயன்படுத்தலாம். அதாவது “அ”வை அழுத்தினால் “அ” சேர்ந்த உயிர்மெய் எழுத்தில் இருந்து “அ” – வை நீக்கி புள்ளியாக “அ” சேர்ந்த உயிர்மெய் எழுத்தின் மேல் வைக்கப்படுவதாகப் பொருள்படுகிறது. மற்ற உயிர் குறியீடுகளை அழுத்தினால், “அ” சேர்ந்த உயிர்மெய் எழுத்தில் இருந்து “அ” – வை நீக்கப்படுவதுடன், அந்தந்த உயிர்க் குறியீடுகள் சேர்க்கப்படுவதாகப் பொருள்படுகிறது. தேவையில்லாத புள்ளி விசை நீக்கப்பட்டுள்ளது.

சொற்பிழைத் திருத்தி (spell checker), இலக்கணப்பிழைத் திருத்தி (grammar checker) ஆகியவற்றை இயக்க மேடையில் ஆங்கிலத்திற்கு இருப்பது போல் மென் பொருளுக்கு வழிகை செய்ய வேண்டும். அதுவரை, சொற்பிழைத் திருத்தி (spell checker), இலக்கணப்பிழைத் திருத்தி (grammar checker) ஆகிய வற்றை “பொது கிடைக்குமிடம்” (open source) போன்ற இணையத்திலிருந்து இலவசமாக இறக்கிப் பயன்படுத்த வகை செய்திட வேண்டும். நடைமுறையில் உள்ள உயிர்க் குறியீடுகள், கைவிடப்பட்ட பழைய எழுத்துக்கள், ஆகியவைகளைச் சின்னங்களாக அதாவது அறிகுறிகளாக (symbols) வைத்துப் பயன்படுத்திட ஒருங்குறி அமைப்புக்கு (Unicode Consortium) எழுதி வகைசெய்திட வேண்டும்.

இணைப்பு: புதிய விசைப்பலகை அமைப்பு மாதிரிகள் (1).கசடதபற /ABCDEF ,(2). அஆஇஈஉஊ / ABCDEF, (3). கசடதபற / QWERT, (4). அஆஇஈஉஊ/ QWERT, (5). தமிழ் 99/QWERT.

1. புதிய மாதிரி விசைப்பலகை அமைப்பு – அஆஇஈஉஊ / ABCDEF

மொழி	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த		
வரிசை எண்	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14
மேற் பக்கம்	தூக்கல்	~			@	@	#	#	[]	[]	%	%	^	^	&	&	*	*	(())	-	-	+	+		\
மேல் வரிசை	சாதா	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஓ	ஔ	ஶ	ஷ	஠	஡	஢	ண	வ	ஶ	ஷ	஠	஡	஢	ண	வ	ஶ	ஷ
நடு வரிசை	தூக்கல்	κ	ஹ	ι	οβF 3	κ	οβF 4	η	οβF 5	ο	οβF 6	ρ	οβF 7	ο	οβF 8	ρ	οβF 9	σ	:	:	^	^	^	^	^	^	^	
கீழ் வரிசை	சாதா	κ	κ	ι	η	η	σ	ρ	ο	ο	ο	ο	ο	ο	ο	ο	ο	ο	:	:	^	^	^	^	^	^	^	
கீழ் வரிசை	தூக்கல்	π	ο	ρ	(ν)	ω	(χ)	γ	:	z	/	<	<	>	>	?	?							
சாதா	κ	ρ	υ	ο	ν	ω	μ	ξ	η	γ	η	z	ο	ξ	

ஆ = ஆங்கிலம், த = தமிழ். மொத்தமுள்ள 94 விசைகள் கீழ்க் காணுமாறு பங்கிடப்பட்டுள்ளன

தமிழ் விசைப்பலகை				
இடம்	வகை	விவரங்கள்	வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	கடைசீ விசை இடப்புறச் சாய்வுக்கோடு அமைந்துள்ளது. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	14
	சாதா	முதல் விசை “ஸ”, கடைசீ விசை “ஷ” அமைந்துள்ளன. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	28
மேல் வரிசை	தூக்கல்	தமிழ் எண்கள் 12 – ம் அமைந்துள்ளன	12	40
	சாதா	உயிர்கள் 12 – ம் அமைந்துள்ளன	12	52
நடு வரிசை	தூக்கல்	“ஹ” கரம், அடுத்து, நாள், மாதம், ஆண்டு, ரூபாய், எண், பற்று, வரவு, அரைப்புள்ளி, ஒற்றை, இரட்டை மேற்கோள் குறியீடுகள் ஆகிய பதினொன்றும் அமைந்துள்ளன	11	63
	சாதா	க,ங,ச,ஞ,ட,ண,த,ந,ப,ம,ய ஆகிய 11 – ம் அமைந்துள்ளன	11	74

கீழ் வரிசை	தூக்கல்	ஆயுதம், அடுத்து, “ப” அடைகுறிகள், “வளைவு” அடைகுறிகள், அரைகாற்புள்ளி, வலப்புறச் சாய்வுக்கோடு, பெரியது சிறியது மற்றும் கேள்விக்குறி ஆகிய பத்தும் அமைந்துள்ளன	10	84
	சாதா	ர,ல,வ,ழ,ள,ற,ன ஆகிய எழும், அடுத்து, காற்புள்ளி, புள்ளி ஆகிய இரண்டும் கடைசியாக “ஐ” வும் ஆக மொத்தம் 10 - ம் அமைந்துள்ளன.	10	94

ஆங்கில விசைப்பலகை				
இடம்	வகை	விவரங்கள்	வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	மாற்றமில்லை	14	14
	சாதா	மாற்றமில்லை	14	28
மேல் வரிசை	தூக்கல்	A முதல் J வரை 10 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	12	40
	சாதா	a முதல் j வரை 10 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	12	52
நடு வரிசை	தூக்கல்	K முதல் S வரை 9 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	11	63
	சாதா	k முதல் s வரை 9 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	11	74
கீழ் வரிசை	தூக்கல்	T முதல் Z வரை 7 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	10	84
	சாதா	t முதல் z வரை 7 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	10	94

ஆ = ஆங்கிலம், த = தமிழ். மொத்தமுள்ள 94 விசைகள் கீழ்க் காணுமாறு பங்கிடப்பட்டுள்ளன

தமிழ் விசைப்பலகை				
இடம்	வகை	விவரங்கள்	வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	கடைசீ விசை இடப்புறச் சாய்வுக்கோடு அமைந்துள்ளது. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	14
	சாதா	முதல் விசை “ல”, கடைசீ விசை “ஷ” அமைந்துள்ளன. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	28
மேல் வரிசை	தூக்கல்	தமிழ் எண்கள் 12 - ம் அமைந்துள்ளன	12	40
	சாதா	நெடில் உயிர்கள் “ஓ, ஏ, ஊ, ஈ, ஆ” வரை ஐந்தும், அடுத்து, மெல்லினம் “ங, ஞ, ண, ம, ன” ஆறும், கடையீயில் “ஜ” - ம் இடம் பெற்றுள்ளன	12	52
நடு வரிசை	தூக்கல்	நாள், மாதம், ஆண்டு, ரூபாய், எண், ” அடுத்து, “ஹ”, அடுத்து, பற்று, வரவு, அரைப்புள்ளி, ஒற்றை, இரட்டை மேற்கோள் குறியீடுகள் ஆகிய பத்தும் அமைந்துள்ளன	11	63
	சாதா	குறில் உயிர்கள் “ஓ, ஏ, உ, இ, அ” வரை ஐந்தும், அடுத்து, வல்லினம் “க, ச, ட, த, ப, ற” ஆறும் அமைந்துள்ளன	11	74

கீழ் வரிசை	தூக்கல்	“ப” அடைகுறிகள், “வளைவு” அடைகுறிகள், அரைகாற்புள்ளி, ஆயுதம், வலப்புறச் சாய்வுக்கோடு, பெரியது சிறியது மற்றும் கேள்விக்குறி ஆகிய பத்தும் அமைந்துள்ளன	10	84
	சாதா	ஒள, ஐ ஆகிய நெடில் இரண்டு, அடுத்து, இடையினம் “ய, ர, ல, வ, ழ” வரை ஐந்தும் அடுத்து,கால் புள்ளி, புள்ளி,ஆகிய இரண்டு குறியீடுகள், கடைசீயாக இடையினம் “ள”வும், அமைந்துள்ளன	10	94

ஆங்கில விசைப்பலகை

இடம்	வகை	விவரங்கள்	வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	மாற்றமில்லை	14	14
	சாதா	மாற்றமில்லை	14	28
மேல் வரிசை	தூக்கல்	A முதல் J வரை 10 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	12	40
	சாதா	a முதல் j வரை 10 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	12	52
நடு வரிசை	தூக்கல்	K முதல் S வரை 9 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	11	63
	சாதா	k முதல் s வரை 9 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	11	74
கீழ் வரிசை	தூக்கல்	T முதல் Z வரை 7 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	10	84
	சாதா	t முதல் z வரை 7 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	10	94

2 புதிய விசைப்பலகை அமைப்பு - ஈடதபறு / ABCDEFJ

மொழி	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந	ஆ	ந									
வரிசை எண்	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	
மேற் பக்கம்	தூக்கல்	-	-	!	!	@	@	#	#	\$	\$	%	%	^	^	&	&	*	*	{	{	}	}	_	_	+	+		
	சாதா	'	ஸ	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	0	0	-	-	=	=	\	ஐ
மேல் வரிசை	தூக்கல்	A	க	B	உ	C	ஊ	D	ச	E	ஐ	F	஋	G	஌	H	அ	இ	ஔ	J	ங	{	{	௩	௩	௪	௪	௫	௫
	சாதா	a	ஔ	b	ர	c	ஊ	d	ஈ	e	ஆ	f	ஊ	g	ஐ	ஊ	h	ஊ	ஊ	ஊ	ஊ	ஊ	{	{	௩	௩	௪	௪	௫
நடு வரிசை	தூக்கல்	K	ஔ	L	ஔ	M	ஔ	N	ஔ	O	ஔ	P	ஔ	Q	ஔ	R	ஔ	S	:	:	'	'	'	'	'	'	'	'	'
	சாதா	k	ஔ	l	ஔ	m	ஔ	n	ஔ	o	ஔ	p	ஔ	q	ஔ	r	ஔ	s	:	:	'	'	'	'	'	'	'	'	'
கீழ் வரிசை	தூக்கல்	T	[U] V	{	W	}	X	:	Y	௩	Z	/	<	<	>	>	?	?									
	சாதா	t	ஔ	u	ஔ	v	ஔ	w	ஔ	x	ஔ	y	ஔ	z	ஔ	/	ஔ								

ஆ = ஆங்கிலம், த = தமிழ். மொத்தமுள்ள 94 விசைகள் கீழ்க் காணுமாறு பங்கிடப்பட்டுள்ளன.

இடம்	வகை	விவரங்கள்	வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	கடைசீ விசை இடப்புறச் சாய்வுக்கோடு அமைந்துள்ளது. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	14
	சாதா	முதல் விசை “ஸ”, கடைசீ விசை “ஷ” அமைந்துள்ளன. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	28
மேல் வரிசை	தூக்கல்	தமிழ் எண்கள் 12 – ம் அமைந்துள்ளன	12	40
	சாதா	உயிர்கள் 12 – ம் அமைந்துள்ளன	12	52

நடு வரிசை	தூக்கல்	“ஹ”கரம், அடுத்து, நாள், மாதம், ஆண்டு, ரூபாய், எண், பற்று, வரவு, அரைப்புள்ளி, ஒற்றை, இரட்டை மேற்கோள் குறியீடுகள் ஆகிய பதினொன்றும் அமைந்துள்ளன	11	63
	சாதா	க,ங,ச,ஞ,ட,ண,த,ந,ப,ம,ய ஆகிய 11 – ம் அமைந்துள்ளன	11	74
கீழ் வரிசை	தூக்கல்	ஆயுதம், அடுத்து, “ப” அடைகுறிகள், “வளைவு” அடைகுறிகள், அரைகாற்புள்ளி, வலப்புறச் சாய்வுக்கோடு, பெரியது சிறியது மற்றும் கேள்விக்குறி ஆகிய பத்தும் அமைந்துள்ளன	10	84
	சாதா	ர,ல,வ,ழ,ள,ற,ன ஆகிய ஏழும், அடுத்து, காற்புள்ளி, புள்ளி ஆகிய இரண்டும் கடைசியாக “ஜ” வும் ஆக மொத்தம் 10 – ம் அமைந்துள்ளன.	10	94

ஆங்கில விசைப்பலகை				
இடம்	வகை	விவரங்கள்	வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	மாற்றமில்லை	14	14
	சாதா	மாற்றமில்லை	14	28
மேல் வரிசை	தூக்கல்	A முதல் J வரை 10 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	12	40
	சாதா	A முதல் J வரை 10 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	12	52
நடு வரிசை	தூக்கல்	K முதல் S வரை 9 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	11	63
	சாதா	K முதல் S வரை 9 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	11	74
கீழ் வரிசை	தூக்கல்	T முதல் Z வரை 7 தலைப்பு எழுத்துக்கள், பிறகு மாற்றமில்லை	10	84
	சாதா	T முதல் Z வரை 7 சாதாரண எழுத்துக்கள், பிறகு மாற்றமில்லை	10	94

1. புதிய மாதிரி விசைப்பலகை அமைப்பு – அஆஇஈஉஊ / QWERT

மொழி	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த											
வரிசை எண்	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	
மேற்பக்கம்	தூக்கல்	~	-			@	@	#	#	\$	\$	%	%	^	^	&	&	*	*	(())	_	_	+	+		
சாதா	ஸ	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	0	0	-	-	=	=	\	ஷ		
மேல் வரிசை	தூக்கல்	Q	க	W	உ	E	ஊ	R	ச	T	ஊ	Y	ஊ	U	எ	I	ஊ	O	ஊ	P	ய	{	எ	}	த				
சாதா	q	ஆ	w	ஆ	e	இ	r	ஈ	t	உ	y	ஊ	u	எ	i	ஏ	o	ஐ	p	ஐ	[ஐ]	ஊ					
நடு வரிசை	தூக்கல்	A	ஹ	S	OB	D	OB	F	OB	G	OB	H	OB	J	OB	K	OB	L	:	:	:	:	:	:	:	:	:	:	:
சாதா	a	க	s	ஊ	d	ச	f	ஊ	g	ட	h	ஊ	j	த	k	ந	l	ப	:	ம	:	:	:	:	:	:	:	:	
கீழ் வரிசை	தூக்கல்	Z	ஃ	X	[C]	V	{	B	}	N	:	M	/	<	<	>	>	?	?								
சாதா	z	ர	x	ல	c	வ	v	ழ	b	எ	n	ற	m	ன	,	/	ஐ								

ஆ = ஆங்கிலம், த = தமிழ். மொத்தமுள்ள 94 விசைகள் கீழ்க் காணுமாறு பங்கிடப்பட்டுள்ளன.

இடம்	வகை	விவரங்கள்	விசைகள்
மேற் பக்கம்	தூக்கல்	கடைசீ விசை இடப்புறச் சாய்வுக்கோடு அமைந்துள்ளது. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14 – 61
	சாதா	முதல் விசை “ஸ”, கடைசீ விசை “ஷ” அமைந்துள்ளன. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14 – 1 4
மேல் வரிசை	தூக்கல்	மேல் வரிசையில் தமிழ் எண்கள் அமைந்துள்ளன	12 – 73
	சாதா	மேல் வரிசையில் உயிர்கள் “ஓ, ஏ, ஊ, ஈ, ஆ,” ஆகிய ஐந்தும் உயிர்களை அடுத்து, “ள,ற,ன,ட,ண,ச,ஞ” ஆகிய ஏழும் அமைந்துள்ளன	12 - 26
நடு வரிசை	தூக்கல்	நடு வரிசையில் நாள்,மாதம், ஆண்டு, ரூபாய், எண், “ஹ”, அரைப்புள்ளி, இரட்டை மேற்கோள் அரைகாற்புள்ளி,ஒற்றை மேற்கோள், வலப்பக்கச் சாய்வுக்கோடு, ஆகிய குறியீடுகள் பத்தும் அமைந்துள்ளன	11 – 84
	சாதா	நடு வரிசையில் உயிர்கள் “ஒ, எ, உ, இ, அ” ஆகிய ஐந்தும் உயிர்களை அடுத்து, “க, ப, ம, த, ந, ய” ஆகிய ஆறும் அமைந்துள்ளன	11 – 37
கீழ் வரிசை	தூக்கல்	கீழ் வரிசையில், பற்று, “ப” அடைகுறிகள், “வளைவு” அடைகுறிகள், ஆயுதம், வரவு, பெரியது சிறியது மற்றும் கேள்விக்குறி ஆகிய பத்தும் அமைந்துள்ளன	10 – 94
	சாதா	கீழ் வரிசையில் உயிர்கள் “ஒள, ஐ” ஆகிய இரண்டு உயிர்களை அடுத்து, “ழ, வ , ங, ல, ர” ஆகிய ஐந்தும் வரிசையில் ரகரத்தினை அடுத்து காற்புள்ளி, புள்ளி, ஆகிய இரண்டு குறியீடுகளும், அதனை அடுத்து கடைசீயாக “ஜ” அமைந்துள்ளன	10 - 47

கருத்துரு – அருள்

(தமிழ் 99 – திருத்திய விசைப்பலகை அமைப்பு)

கருத்துரு – அருள் - ஐ ஏற்று, லைனக்ஸ், மைக்ரோசாப்டு, மேகின்டோஷ், போன்ற அனைத்து கணினி இயக்க மேடைகளிலும் (All Computer Operating Systems like Linux, Microsoft, Macintosh), டாலி, எம்.எஸ் ஆபீஸ், ஓபன் ஆபீஸ், ஸ்டார் ஆபீஸ் போன்ற பயன்பாட்டு மென்பொருட்களிலும் (in all application softwares) தானமைவாக நிறுவிட (installed as default/Inscript) இந்திய அரசு, ISI துறை, ஒப்புதல் வழங்க தொடர் நடவடிக்கை எடுக்க பரிந்துரை செய்ய வேண்டுமென்று பணிவன்புடன் கேட்டுக்கொள்கிறேன். இந்திய அரசு, ISI துறை, ஒப்புதல் கொடுத்த விசைப்பலகையோடு யூனிகோடு உருவேற்றம் (Unicode encoding), யூனிகோடு உருவிறக்கம் (Unicode decoding) மற்றும் யூனிகோடு எழுத்துரு (Unicode fonts) ஆகியவை இணைவதால் கணினித்தல் தரவுப் பயன்பாடு, மின்னஞ்சல் இணைய பயன்பாடு முதலியவைகளை உலகளவில் சிக்கலின்றி அனைத்துத் தரப்பினரும் பயன்படுத்த வழிவகை ஏற்படும் என்று பணிவன்புடன் தெரிவித்துக்கொள்கிறேன்.

குறிப்பு: 1. அகரம் சேர்ந்த உயிர்மெய்எழுத்துக்களுக்கு “அ” விசை தேவைப்படவில்லை. அதனால் அந்த “அ” (ஐ) விசையை அனைத்து மெய் எழுத்துக்களுக்கு புள்ளி இடுவதற்குப் பயன்படுத்தலாம். அதாவது “அ”வை அழுத்தினால் “அ” சேர்ந்த உயிர் மெய் எழுத்தில் இருந்து “அ” நீக்கப்பட்டு புள்ளி இடப்படுவதாகப் பொருள்படுகிறது. மற்ற உயிர்க்குறியீடுகளை அழுத்தினால், “அ” நீக்கப்பட்டு, அந்தந்த உயிர்கள் சேர்க்கப்படுவதாகப் பொருள்படுகிறது. அதனால், தமிழ் 99 – ல் உள்ள, புள்ளி விசை நீக்கப்பட்டுள்ளது.

தமிழில் கணினித்தல் எளிதாக அமைய கீழ்க்காணும் வசதிகளையும் செய்துதர நடவடிக்கை எடுக்க வேண்டும்.

(1) சொற்பிழைத் திருத்தி (spell checker), இலக்கணப்பிழைத் திருத்தி (grammar checker) ஆகியவற்றை இயக்க மேடையில் ஆங்கிலத்திற்கு இருப்பது போல் தமிழுக்கும் வழிகை செய்ய வேண்டும். அதுவரை, சொற்பிழைத் திருத்தி (spell checker), இலக்கணப்பிழைத் திருத்தி (grammar checker) ஆகியவற்றை “பொது கிடைக்குமிடம்” (open source) போன்ற இணையத்திலிருந்து இலவசமாக இறக்கிப் பயன்படுத்த வகை செய்திட வேண்டும்.

(2) தமிழில் கூட்டொலி எழுத்துக்களுக்கு இடமில்லை. அதனால் சஷகரம், ஸ்ரீகாரம் இரண்டும் கூட்டொலி எழுத்தாக இருப்பதால் தமிழ் நெடுங்கணக்கு வரிசையிலிருந்து அகற்றி அவற்றைச் சின்னங்களாக அமைத்துப் பயன்படுத்திட ஒருங்குறி அமைப்புக்கு (Unicode Consortium), எழுதி வகைசெய்திட வேண்டும். எழுத்துச் சீர்திருத்தம் காரணமாக புதிய உயிர்குறியீடுகள், நடைமுறையில் உள்ள உயிர்க் குறியீடுகள், கைவிடப்பட்ட பழைய எழுத்துக்கள் ஆகிய இவைகளைச் சின்னங்களாக அதாவது அறிகுறிகளாக (symbols) வைத்துப் பயன்படுத்திட ஒருங்குறி அமைப்புக்கு (Unicode Consortium) எழுதி வகைசெய்திட வேண்டும்.

இணைப்பு: (1) கருத்துரு - அருள், விசைப்பலகை அமைப்பு

கருத்துரு - அருள் (தமிழ் 99 - திருத்திய விசைப்பலகை அமைப்பு)

மொழி	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த	ஆ	த				
வரிசை எண்	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14		
மேற்பக்கம்	தூக்கல்	~	~			@	@	#	#	\$	\$	%	%	^	^	&	&	*	*	(())	-	-	+	+		\	
சாதா	சாதா	.	ஸ	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	0	0	-	-	=	=	\	ஷ	
மேல் வரிசை	தூக்கல்	Q	க	W	உ	E	ஊ	R	ச	T	டு	Y	சூ	U	எ	I	அ	O	க	P	ய	{	ள	}	கூ					
சாதா	சாதா	q	ஓ	w	ஏ	e	ஊ	r	ஈ	t	ஆ	y	ள	u	ற	i	ன	o	ட	p	ண	[ச]	ஞ					
கீழ் வரிசை	தூக்கல்	A	OBF3	S	OBF4	D	OBF5	F	OBF6	G	OBF7	H	ஹ	J	;	K	“	L	:	:	‘	“	/							
சாதா	சாதா	a	ஓ	s	எ	d	உ	f	இ	g	அ	h	க	j	ப	k	ம	l	த	;	ந	‘	ய							
கீழ் வரிசை	தூக்கல்	Z	OBF8	X	[C]	V	{	B	}	N	ஃ	M	<	<	>	>	?	?										
சாதா	சாதா	z	ஓ	x	ஐ	c	ழ	v	வ	b	ங	n	ல	m	ர	,	,	.	.	/	ஜ									

ஆ = ஆங்கிலம், த = தமிழ். மொத்தமுள்ள 94 விசைகள் கீழ்க் காணுமாறு பங்கிடப்பட்டுள்ளன.

இடம்	வகை	விவரங்கள்	விசைகள்	
			வரிசை	கடுதல்
மேற்பக்கம்	தூக்கல்	கடைசீ விசையில் மட்டும் இடப்புறச் சாய்வுக்கோடு அமைந்துள்ளது. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	14
	சாதா	முதல் விசையில் “ஸ”, கடைசீ விசையில் “ஷ” அமைந்துள்ளன. மற்றவை ஆங்கில விசைப்பலகையில் உள்ளது போல் கணக்கியல் குறியீடுகள் அமைந்துள்ளன	14	28
மேல் வரிசை	தூக்கல்	தமிழ் எண்கள் 12 - ம் அமைந்துள்ளன.	12	40
	சாதா	ஐந்து விசைகளில் உயிர்கள் “ஓ, ஏ, ஊ, ஈ, ஆ,” ஆகிய ஐந்தும், அடுத்து, “ள,ற,ன,ட,ண,ச,ஞ” ஆகிய உயிர்மெய் எழுத்துக்கள் ஏழும் அமைந்துள்ளன	12	52

நடு வரிசை	தூக்கல்	ஐந்து விசைகளில் நாள்,மாதம், ஆண்டு, ரூபாய், எண் குறியீடுகளும் நடுவில் “ஹ”, அடுத்த ஐந்து விசைகளில் அரைப்புள்ளி, இரட்டை மேற்கோள் அரைகாற்புள்ளி, ஒற்றை மேற்கோள், வலப்பக்கச் சாய்வுக்கோடு குறியீடுகளும் அமைந்துள்ளன	11	63
	சாதா	ஐந்து விசைகளில் உயிர்கள் “ஒ, எ, உ, இ, அ” ஆகிய ஐந்தும் அடுத்து, “க, ப, ம, த, ந, ய” ஆகிய உயிர்மெய் எழுத்துக்கள் ஆறும் அமைந்துள்ளன	11	74
கீழ் வரிசை	தூக்கல்	பற்று, “ப” அடைகுறிகள், “வளைவு” அடைகுறிகள், ஆயுதம், வரவு, பெரியது சிறியது மற்றும் கேள்விக்குறி ஆகிய பத்தும் அமைந்துள்ளன	10	84
	சாதா	இரண்டு விசைகளில் உயிர்கள் “ஒள, ஐ” அடுத்து, “ழ, வ, ங, ல, ர” ஆகிய ஐந்து உயிர்மெய் எழுத்துக்கள், அடுத்து காற்புள்ளி, புள்ளி, ஆகிய இரண்டு குறியீடுகள், அதனை அடுத்து கடைசீயாக “ஜ” அமைந்துள்ளன	10	94

A Free Tamil Keyboard Interface for Business and Personal Use

Panmozhi Vaayil - A Multilingual Indic Keyboard Interface

Abhinava Shivakumar, Akshay Rao, Arun S, A. G. Ramakrishnan

MILE Lab, Department of Electrical Engineering, Indian Institute of Science

Abhinav.zoso, u.akshay, sarun87, agrkrish@gmail.com

Abstract

A multilingual indic keyboard interface is an Input Method [1] that can be used to input text in various Indic languages like Tamil, Kannada, Telugu, Hindi, Gujarati, Marathi and Bengali. The input can follow the phonetic style making use of the standard QWERTY layout along with support for popular keyboard and typewriter layouts [1] (also known as soft layouts) of Indic languages using overlays. Indic-keyboards provides a simple and clean interface supporting multiple languages and multiple styles of input working on multiple platforms. XML based processing makes it possible to add new layouts or new languages on the fly. These features, along with the provision to change key maps in real time make this input method suitable for most, if not all text editing purposes. Since Unicode is used to represent text, the input method works with most applications. This is available for free download and free use by individuals or commercial organizations, on code.google.com under Apache 2.0 license.

Keywords-Input Method, IM, Indic, Localization, Internationalization (i18n), FOSS, Unicode, panmozhi vaayil, vishwa vaangmukha

Introduction

Input method editors or IMEs provide a way in which text can be input in a desired language. Traditionally, IMEs are used to input text in a language other than English. Latin based languages (English, German, French, Spanish, etc.) are represented by the combination of a limited set of characters. Because this set is relatively small, most languages have a one-to-one correspondence of a single character in the set to a given key on a keyboard. When it comes to East Asian languages (Chinese, Japanese, Korean, Vietnamese etc.) and Indic languages (Tamil, Hindi, Kannada, Bangla etc.), the number of key strokes to represent an akshara can be more than one, which makes using one-to-one character to key mapping impractical. To allow for users to input these characters, several input methods have been devised to create Input Method Editors. The term *input method* generally refers to a particular way to use the keyboard to input a particular language. The term *input method editor* refers to the actual program that allows an input method to be used.

Objective

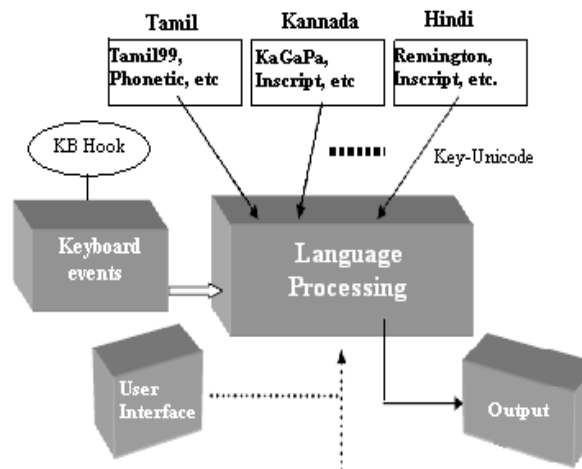
The focus has been to develop a multilingual input method editor for Indic languages. The interface should be minimalistic in nature providing options to configure and select various language layouts. Configurability is inclusive of addition of new layouts or languages and option to enable or disable the . Inputs can be based on popular keyboard layouts or using a phonetic style [2]. We call it Panmozhi

Vaayil in Tamil and Vishwa Vaangmukha: in Sanskrit, both meaning entrance for many languages. It is known by the generic name, Indic Keyboard IM in English and is available for download from <http://code.google.com/p/indic-keyboards/>

Motivation

Some of the main reasons for developing indic-keyboards are as follows:

1. To ease inputting of any Indian language under any platform.
2. To facilitate increased use and presence of Indian languages on the computer and internet.
3. To provide a free with an unrestricted license.



4. Phonetic as well as popular layout support in a single package.
5. Need for a unified multiplatform input method.
6. Ease of configurability and customizability.

Figure 1 shows the system architecture of Panmozhi Vaayil showing the various modules and their interaction.

Existing works and what they offer

A good amount of effort has already gone into the development of easy, flexible input methods. Some popular ones are:

- **Baraha IME** – Provides phonetic support for a fixed number of languages and designed for use on Microsoft Windows platform [3].
- **Aksharamala** – Similar to BarahaIME with support for Microsoft Windows [4].
- **Smart Common Input Method (SCIM)** – Designed to work on Linux with phonetic style of input [5].

What indic-keyboards (Panmozhi Vaayil) offers

- No installation hassles.
- Phonetic as well as popular keyboard layouts.
- Dynamic module enabling the addition of new keyboard layouts by even users.
- Both on Linux platform and Microsoft Windows.
- Phonetic key maps can be changed to meet user's requirements.

- Open source.
- Available under Apache 2.0 License, which means even commercial companies can use our code to develop products, after acknowledging us.

Design

The design can be broadly categorized into the following modules:

- User interface and the shell extension.
- Capturing the keyboard events.
- XML based Unicode processing.
- Rendering the Unicode.

The following block diagram shows the architecture:

User Interface and Shell Extension

The User interface is a shell extension which sits in the system tray/Notification area. The main purpose of this is to allow users to interact with the input method. This mainly involves selection of the language and the particular keyboard layout. It also helps in enabling and disabling the input method, accessing help and to display the image of the keyboard layout currently selected. Apart from these, the menu also has provision for addition of new keyboard layouts.

Capturing the Keyboard Events

The input method is designed to operate globally. That is, once the input method is enabled, further key strokes will result in characters of the particular language selected being rendered system wide. This requires the capture of the key presses system wide across all processes. A keyboard hook installed in the kernel space will enable this. This module is, therefore, platform specific.

XML Based Unicode Processing

Finite Automata exists for each language (for every keyboard layout). The finite automata has been designed as XML files, where every XML file corresponds to a keyboard layout. XML based processing makes it possible to add new layouts or new languages dynamically. The XML file has a pattern, which corresponds to the input key(s) pressed. The input pattern is matched and required processing is carried out to see if the pattern matched is a vowel or a consonant. For the input pattern, a sequence of Unicode(s) is returned. The structure of the XML file is as shown below:

```
<pattern>
  <char>A</char>
  <unicode>0C86</unicode>
  <consonant>0</consonant>
  <uni2>0CBE</uni2>
</pattern>
```

The above XML block indicates that for the key press "A", the corresponding Unicode is 0C86. The consonant tag tells us whether the key is a vowel or a consonant. In case it is a vowel, a second tag indicates the Unicode of the associated dependent vowel (if any).

Two algorithms have been designed, one for phonetic style of input and the other for keyboard layouts. Both the algorithms are generic, i.e. same algorithm is used for keyboard layouts of all languages and one algorithm for phonetic input in any language.

The XML key maps can be changed on the fly and the changes are reflected instantly.

Unicode Rendering.

Once the key is pressed, simple grammar rules are applied to determine whether the output has to be a consonant, an independent vowel or a dependent vowel. The XML file is parsed and the corresponding Unicode is fetched. The Unicode is sent back to the process, where the keypress event took place and is rendered if any editable text area is present. The rendering of Unicode is platform specific.

Implementation

The following tools and languages have been used to implement the input method:

Java SE - The main language processing module has been implemented using Java SE. This has enabled easy portability and up to 80% of the code has remained common across platforms.

Eclipse SWT - Used to implement the user interface. Eclipse SWT uses Java SWT is preferred over other toolkits to get a native look and feel..

XML - As described previously, finite automata exists for every language (layout) and XML has been used to design it. Simple API for XML Parsing (SAX) has been used to parse the XML.

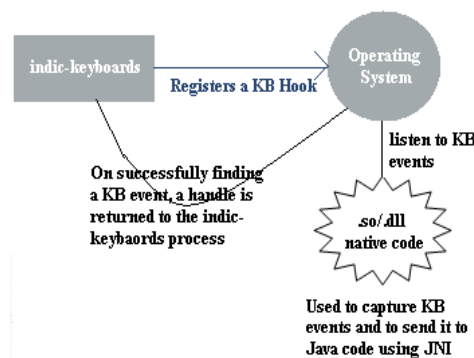
Win32 libraries : The Windows API, is Microsoft's core set of application programming interfaces (APIs) available for the Microsoft Windows platform. Almost all Windows programs interact with the Windows API. Some examples are SAPI, Tablet PC SDK, Microsoft Surface etc. Platform specific portions of the input method has been implemented to run on Microsoft Windows variants using the Microsoft Win32 libraries. Both keystroke capturing and Unicode rendering have been accomplished using Win32 libraries.

Steps :

a. Syshook.dll : Install a keyboard hook in the operating system. The hook is set up for the keyboard to listen to key presses. The Windows API used is **SetWindowsHookEx()** and the library accessed is user32.dll. (See Fig. 2)

b. opChars.dll : Responsible for putting the character on to the current active window. Sends a message to the input event queue using the Windows API **SendInput()**. The library accessed is user32.dll

Fig. 2. JNI-Native code and keyboard hook procedure



Evdev - Also known as the user input device event interface/Linux USB subsystem. Used to capture keystrokes in GNU/Linux. This provides a clean interface to handle input devices like keyboard, mouse, joystick etc. in the userspace in Linux. This involves the following things :

- a. Open file descriptor to the input device using the **open()** API with suitable parameters. Use **ioctl()** to identify the device.
- b. Using the open file descriptor, continuously read bytes from the input device (can be key press/release, mouse movements) using the **read()** API

Xlib - Used for Unicode rendering in GNU/Linux.

Steps :

- a. Identify the current active window using **XGetInputFocus()**
- b. Make the window listen to all keypress events using **XSelectInput()**
- c. Using the keycodes obtained for every keypress/release event from evdev, using a mapping table to map the keycode to the keysym. Output the Unicode to the active window using **XSendEvent()** API

Java Native Interface - Also known as JNI in short. The JNI enables the integration of code written in the Java programming language with code written in other languages such as C and C++. The write once, run anywhere concept arises from the fact that Java acts as an abstraction layer on top of the native implementation. All the API java provides have been natively implemented and the Java code allows the same APIs to be used across platforms.

The native code is usually packaged as a DLL or a Shared Object. The Java method which accesses the native code is created with a keyword "native". Header files need to be created for the classes which contain these methods. At run-time, java code interacts with the native libraries using predefined interfaces. The native methods can also call Java methods. This mechanism is known as **JNI callback**.

Languages and Layouts Supported

TABLE I - LANGUAGES AND LAYOUTS SUPPORTED

Language	Phonetic	Layouts
Hindi	Yes	Inscript Remington
Kannada	Yes	Inscript KaGaPa
Tamil	Yes	Inscript Tamil99 Remington
Telugu	Yes	Inscript
Gujarati	Yes	Inscript
Marathi	Yes	Inscript Remington
Bengali	No	Inscript
Malayalam	No	Inscript
Oriya	No	Inscript
Gurmukhi	No	Inscript

Currently, the input method supports 10 Indian languages namely Tamil, Telugu, Kannada, Malayalam, Hindi, Bengali, Gurmukhi, Oriya, Gujarati and Marathi. The different keyboard layouts currently supported in these languages are listed in Table I.

An easy-to-use user interface has been provided to add new layouts which are Inscript like. Additional phonetic or other layouts can be added based on the existing layouts by creating new XML files and following the prescribed structure. Existing layouts can be changed/customized to suit the users' needs. In phonetic layouts, a single key press to vowel mapping is used to ensure lesser key presses for the completion of the CV combination. Ex : k (க) + Y (ய) = கய instead of k + ae/ai.

Performance

The input method is multithreaded and the following runtime statistics have been obtained. Java Monitoring and Management console has been used to profile.

- Average Heap Memory usage : 4.0 MB (maximum : 5.0 MB)
- CPU usage : 0.2% - 0.3%
- Garbage Collector: Average time for one sweep - 0.05s. Average heap space freed up - 1 MB
- Number of threads: Peak - 15. Average live threads - 13 (2 threads are spawned by the input method)

Conclusion

In conclusion, the project has been an attempt to providing a very dynamic input method editor. The flexibility of adding new Indic languages on the fly, modification of the existing layouts, changing the keypress - Unicode input combination for phonetic input and a host of many other features makes for very easy to use . The main focus has been on flexibility; ease of use and to keep things to a minimum. Keeping that in mind, we have abstained from touching or modifying any system files as well as relieving the user of all installation hassles. Performance wise, the input method is fast and very light on system resources. From the user's perspective, all one needs to do is run it. This also means the user can run the through a pen drive, CD, DVD, hard disk or any portable media. The being open source and licensed under the Apache 2.0 License, developers and users alike can modify, recompile, or rewrite the entire source and can also make these appendages closed source. Apache 2.0 license also allows developers to sell the modified code. All in all, a very dynamic, flexible, easy to use, clean, unrestrictive input method has been designed, which is multiplatform and multilingual.

References

- [1] Russ Rolfe. "What is an IME (Input Method Editor) and how do I use it?", Microsoft Global Development and Computing Portal, July 15, 2003
- [2] A.G. Ramakrishnan, Abhinava Shivakumar, Akshay Rao, Arun S. "Indic-Keyboards - A Multilingual Indic Keyboard Interface", TIC 2009 Conference Book, p. 130-132, INFITT 2009.
- [3] Baraha - Free Indian Language (<http://www.baraha.com>)
- [4] Aksharamala (<http://www.aksharamala.com/>)
- [5] *Smart Common Input Method (SCIM)* (<http://www.scim-im.org/>)

Authors

Abhinava Shivakumar, Project Staff, MILE Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore. abhinav@mile.ee.iisc.ernet.in

Akshay Rao, Project Staff, MILE Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore. u.akshay@gmail.com

Arun S., Project Staff, MILE Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore. sarun87@gmail.com

A. G. Ramakrishnan, Professor, Department of Electrical Engineering, Indian Institute of Science, Bangalore. ramkiag@ee.iisc.ernet.in

Development and Evolution of Tamil Keyboard Input Systems

Ravindran K. Paul

Malaysia (ravi@tsc2.com)

There have been major developments in design of Tamil Keyboard input systems over the past 30 years. A gradual evolution has taken place in the available input systems with changes in computer technology and Operating Systems. Other than the input designs there have also been changes in the philosophies behind the input systems.

This paper will trace the evolution of the input methods and philosophies in line with the changes in computer technology. There were several significant evolutionary steps that took place in this development. These little evolutionary steps were only possible because of the contributions of several individuals and organizations.

This overview is based on personal observations and involvement in keyboard design over the past 25 years. This is not a historical record, but an overview based on information and personal interaction with key players in this field. As such other developments that I have had not access to are not included in this paper.

This evolution in keyboard design was only made possible by the contributions of several key individuals. This overview will show that some of the current thinking on Tamil Keyboard input systems are clearly outdated and need to be changed to take into account current technological and social developments.

Brief History

Tamil Computer software first started to appear in the mid 1980's in DOS based computers. During this time, Tamil software had to be written by programmers as all Tamil characters need to be displayed and printed by the program. There were very few keyboard layouts and programs during this period.

In the late 1980's Windows based software began to appear. This allowed non-programmers to create Tamil fonts to type Tamil. This introduced more keyboard designs as more people were able to create Tamil input systems.

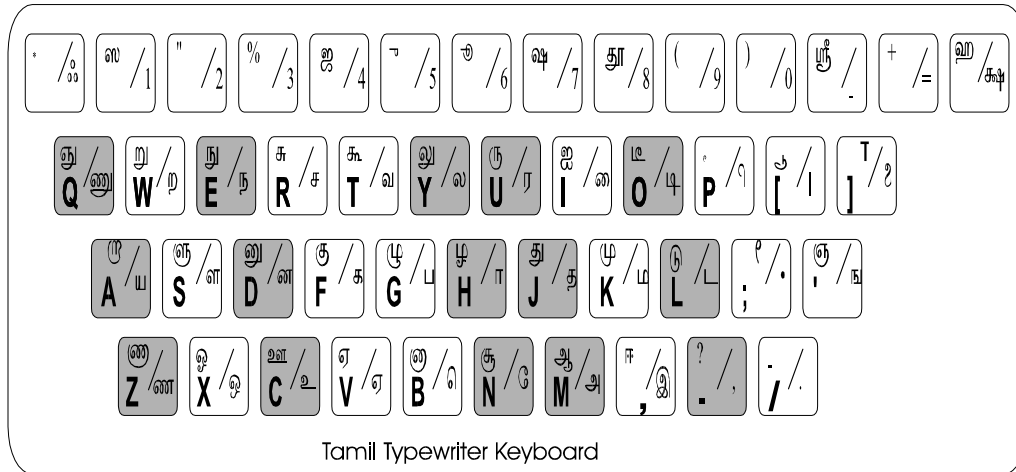
In the mid 1990's the internet revolution started and international discussion on Keyboard designs began. Based on these discussions and the Tamil Nadu Government became involved. This led to the introduction of the Tamil99 keyboard.

In the late 1990's Romanised keyboards began to emerge. Software developers like Murasu and Tamizha began to create keyboards that allowed Tamil typing using the English keyboard.

Currently all these keyboard designs are available in many software. This development of Tamil keyboard layouts over the past 25 years has seen great changes in the philosophy of keyboard design. These changes will be looked to in detail in the next few pages.

Tamil Typewriter Keyboard

In the early 1980's the most popular Tamil keyboard layout was the "Remington Typewriter" layout. The typewriter was the only way to type Tamil text and this was the keyboard layout used by journalists and writers. The first few Tamil software that were released used this design. The buyers of these software were mainly professionals and the software followed the design that they were familiar with.



தமிழ் :	J	=>	த	=	த
	P+K	=>	ி + ம	=	மி
	; +Shift H	=>	ீ + ஸ்ரீ	=	ஸ்ரீ
தொ :	B+J+H	=>	ெ+த+ஈ	=	தொ

In the keyboard each character is a "glyph" or picture. The user typed according to the visual appearance of the characters. In other words the user typed what the user expected to see. It must be noted that because of the mechanical limitations of the type writer :

மி = ி + ம and not மி = ம + ி

In conclusion, the first keyboard input systems were visually based. The user typed the "image" that was to appear. Not much thinking was involved.

Phonetic Keyboard.

In 1986, Thunaivan introduced the worlds first Tamil Phonetic Keyboard. This was a totally new concept at the time. For the first time what the typist typed and what appeared on the screen was not the same.

A Phonetic Keyboard was defined as follows :

The phonetic keyboard is based on the use of only 13 keys for vowels and 17 keys for the consonants. With just these 30 keys, all the keys in the Tamil alphabet can be typed.

*Note : This is a personal definition of the term coined in 1986. Whether the term is correct or not is not the subject of discussion. I have seen other definitions for this term.

The principle is as follows :

uvir elutthu - uyir elutthu
 ு = ு
 ூ = ூ

mei elutthu - mei elutthu
 க = க
 ட = ட

(Note that the definition of "mei elutthu" has been modified from க to க)

<u>mei elutthu</u>	-	<u>uyir elutthu</u>	-	<u>uyirmei elutthu</u>
க	+	஁	=	கீ
ட	+	ஂ	=	ஃ
க	+	ஃ	=	கை
க	+	஄	=	கை

Because there were less keys to press, it was easier to memorise and type in Tamil. Due to this new innovation, more and more teachers started to type in Tamil. Due to my limited knowledge in Tamil, the



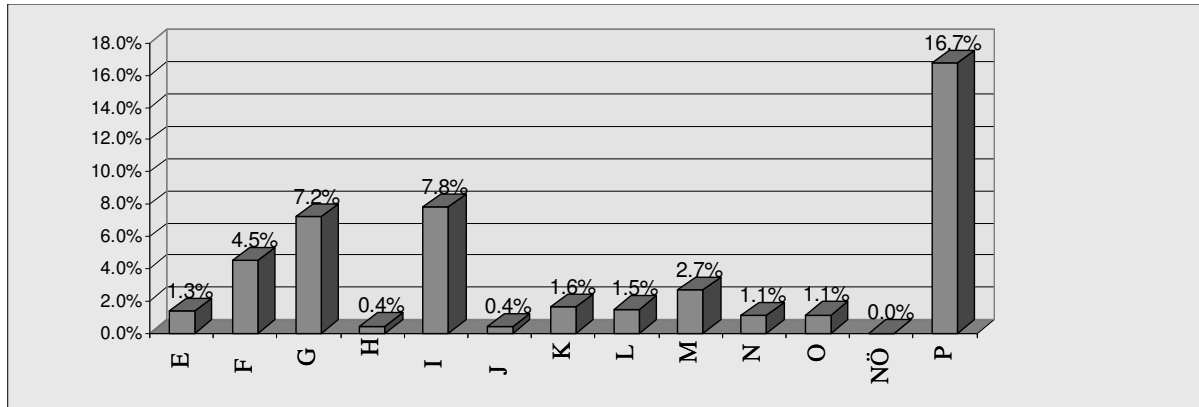
keyboard layout was quite inefficient.

World's first Tamil Phonetic Keyboard.

Refinements to Phonetic Design

In the first Tamil Phonetic keyboard, the arrangement of the keys was not efficient. The next development was the repositioning of keys for optimum typing speeds. In 1987 based on discussions with the late Naa Govindasamy of the Institute of Education in Singapore, both of us decided to look for a more efficient keyboard layout. This resulted in the Thunaivan Advanced Phonetic Keyboard and the I.E. Singapore Keyboard.

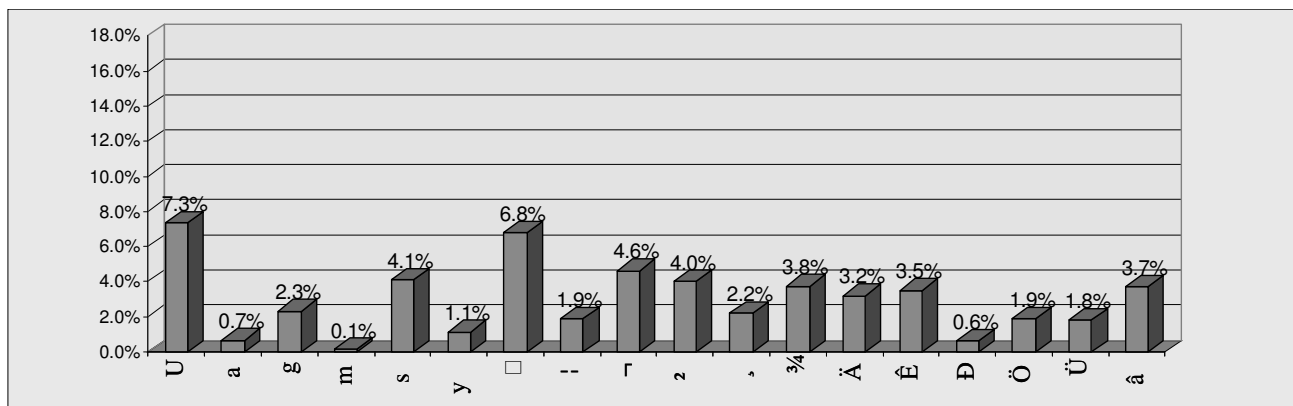
These designs were based on the frequency analysis of keys needed to type Tamil. Following were the results of key use analysis :



Bar chart on frequency of Vowel Column

Note the relatively high frequency of the vowel keystrokes over consonants below.

Bar chart on frequency of Consonant Column

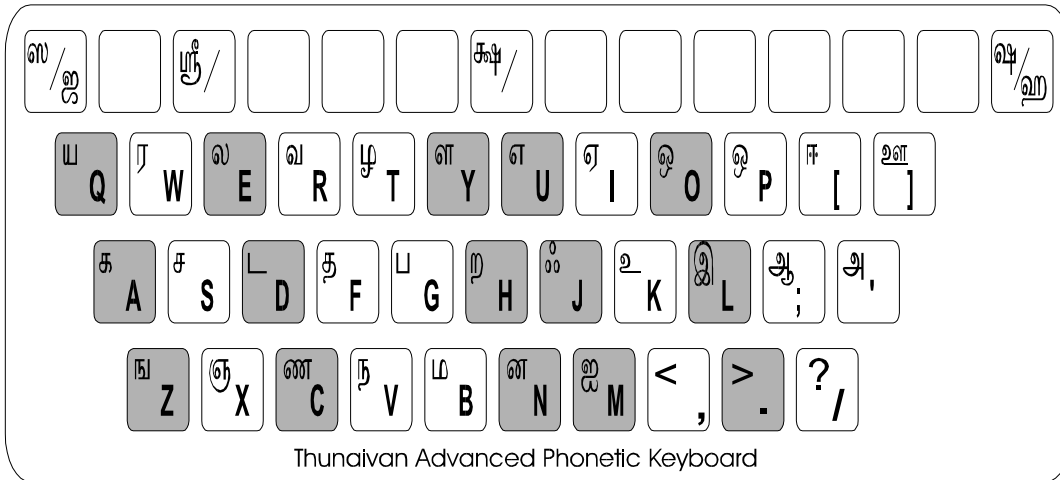


The 3 most frequent consonants in use are ச , த and ட . The next few characters varied in frequency depending on type of document. After this, the most frequent characters varied between ர , ம and ல .

The highest frequency for a consonant is 7.3% for ச , while the highest frequency for the vowel ஊ is more than double at 16.7%. Similarly இ and உ have relatively high frequencies of 7.2% and 7.8% respectively. It would be very important to make sure that these 3 consonants and these 3 vowels are on the home keys, especially ஊ .

Although the objective was to create the most efficient layout possible, the above layout was chosen to help faster key memorizations. Also characters are from left to write. It was assumed that since we write from left to write, it should be more natural to type from left to write.

Despite this compromise, the keyboard was surprisingly efficient. 56% of all frequently used Tamil words can be typed with just the HOME keys (ASDF and JKL;)

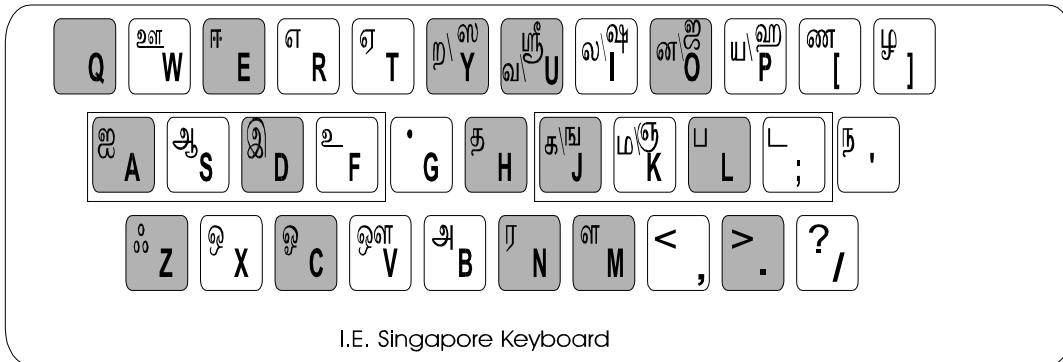


Thunaivan Phonetic Keyboard Keystroke Distribution

Percentage of Keystrokes per Finger, excluding Home Keys	6.1%	4.9%	5.8%	26.3%								Total Home Keys	56.8%	
	2.9%	3.9%	4.3%	14.7%		11.6%	1.5%	1.1%	3.3%					
	Q	W	E	R	T	Y	U	I	O	P	[]		
	2.2%	3.8%	3.2%	3.5%	0.6%	1.9%	1.6%	1.5%	1.1%	1.1%	0.4%	0.4%		
Left Home Keys	Home	A	S	D	F	G	H	J	K	L	;	'	Right Home Keys	
	20.6%	7.3%	2.3%	4.1%	6.8%	4.6%	1.8%	16.7%	7.8%	7.2%	4.5%	1.3%	36.3%	
	Z	X	C	V	B	N	M	,	.	/				
% Mei	53.7%	0.7%	0.1%	1.1%	1.9%	4.0%	3.7%	2.7%					% Uyir	46.3%

I.E. Singapore Keyboard

The I.E. Singapore keyboard worked on a slightly different philosophy. The primary goal was for optimum keyboard placement. Secondly, the keystroke was right to left. The assumption was that the right hand was stronger and should be faster. Also, some ideas were borrowed based on the Phonetic system in the Hindi language Keyboard developed by Mohan Thambi in India in 1983.



Unfortunately there was a mistake in the design. The "pulli" which should have been placed on the "F" key was instead placed on the "G" key. This was because the author had wrongly assumed that the HOME keys were "SDFG" and "JKL;". This was corrected in other keyboard layouts that were introduced later like the Nalinam, Tamil97 and Tamil99 keyboard layouts.

Influence of Microsoft Windows

The above developments are all pre-Windows. All these ideas were implemented in DOS based computers. With the introduction of Windows more keyboard layouts started to come out. Many of these new keyboard layouts were font based.

The first few fonts designed for Windows followed the Typewriter Keyboard layout, but with one significant difference.

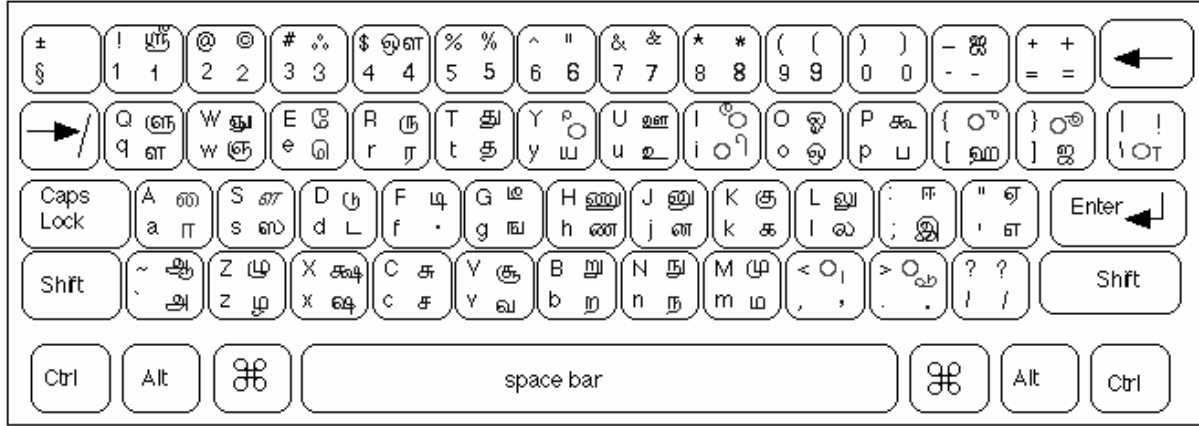
$$\text{மி} = \text{ம} + \text{ி} \quad \text{and not} \quad \text{மி} = \text{ி} + \text{ம}$$

Basically this was the Tamil Typewriter keyboard in reverse. For the first time, non-programmers started to create Tamil input systems. By creating Tamil fonts they were able to type in Tamil. Popular fonts like Tharagai, Baamini and Mylai fonts came into existence.

Romanised Fonts

From these many fonts, one font took a slightly different approach and became very popular. This was the Mylai font by Dr. K.Kalyanasundaram. This popular font tried to match Tamil with English sounds. As shown below most of the Tamil characters were mapped to the English Alphabet.

Mylai Keymap Layout for use with TSCII Tamil font



Note: modifiers are indicated in the red (incl. of au-kara one placed at S). uyir 'au' is at \$ (shift-4).
 New: double quote is at ^ (shift-6) ; exclamation mark at | (shift 1) ; copyright sign © at @ (shift-2)

All the developments above are pre-1995. With the start of online discussions, the need for standardisation became more important. After the first Tamil Internet conference held in Singapore in 1997 discussions on a common keyboard design began. In 1999, the Tamil Nadu government released the Tamil99 Keyboard layout as well as the TAM and TAB font encoding.

Phonetic Keyboard enhancements of Tamil 99

While previous phonetic keyboard designs had concentrated on optimum key locations, the Tamil 99 keyboard took it one step further. One of the design goals was to decrease the number of keystrokes. The Tamil 99 keyboard achieved this in the following way.

Normal Phonetic Keyboard

தொடக்க :	த+ஒ	=	தொ	ங + க	=	ங்க
	ட	=	ட	ஞ + ச	=	ஞ்ச
	க+ஃ	=	க்	ண + ட	=	ண்ட
	க	=	க	ந + த	=	ந்த
Tamil99 Phonetic Keyboard				ம + ப	=	ம்ப
தொடக்க :	த+ஒ	=	தொ	ன + ற	=	ன்ற
	ட	=	ட			
	க+க	=	க்க			

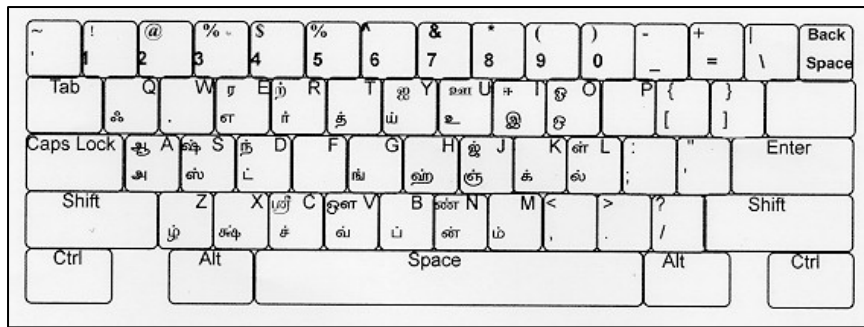
Note the use of க+க above. This function allows the user to save keystrokes, by not having to type an extra key for the "pulli". If there is a need for the sequence கக type the sequence க+அ+க. This applies to all consonants and is designed to save keystrokes.

This combination technique also applies to the following characters :

This new enhancement managed to reduce the number of keystrokes required and made the keyboard layout more efficient. Some of these techniques were adapted for use in later Tamil Romanized Keyboards.

Romanized Keyboards

Around this time Romanized keyboards began to appear. One of the first implementations of this keyboard method was developed by Muthu Nedumaran of Malaysia as part of Murasu Anjal. Other than typing directly in English, there is another interesting change here. This is the use of "mei" (ஃ) instead of "uyirmei" (ஔ) for consonant key strokes. For example typing the "K" key produced "ஃ" instead of "ஔ" as was normal in previous Keyboard Layouts. This was a significant change.

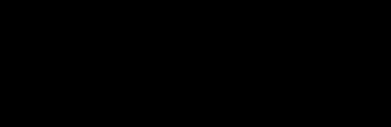


The Romanized keyboard is the last of the innovations in Tamil Keyboard development. It incorporates all the innovations that have taken place in the past 20 years. This keyboard design has had a major impact in the number of people typing Tamil. For the first time, users could start typing in Tamil at reasonable speeds within 20 minutes. Speeds that took 2 to 3 weeks to achieve could now be attained in an hour or less.

There is still some debate on which keyboard is the best or which is the best keyboard should be taught to children. Currently adults and children learn the English keyboard first when they start using computers. Younger children and teenagers reach significant typing speeds English because of high Internet use. Considering the above and the increase use of Tamil as a second language and not as first language in previous generations, the Romanized Tamil Keyboard is very suitable to be taught to adults and children alike.

11

தமிழ் வலைப் பூக்கள்



தமிழ் வளர்ச்சியில் வலைப்பூக்கள்

முனைவர். துரை. மணிகண்டன்.

உதவிப் பேராசியர், தமிழ்த்துறை,
டாக்டர் கலைஞர் கலை அறிவியல் கல்லூரி

(பாரதிதாசன் பல்கலைக்கழக கல்லூரி)

இலால்குடி, திருச்சிராப்பள்ளி.

மின்னஞ்சல்: mkduraimani@gmail.com

முன்னுரை

21ஆம் நூற்றாண்டு அறிவியல் வளர்ச்சியில் அசைக்கமுடியாத இடத்தைப் பெற்றிருப்பது இணையமாகும். தகவல் தொழில் நுட்ப உலகில் இணையம் மிகப்பெரிய உதவிகளை மொழி, இனம், பாராமல் உலக மக்களுக்குச் செய்து வருகின்றது. இது விஞ்ஞானம், அறிவியல், கணக்குகள் என்ற ஒரு குறிப்பிட்ட சிலவற்றிற்குப் பயன்படாமல் இலக்கிய வளர்ச்சிக்குப் பெரிதும் பங்காற்றி வருகின்றது. நெடிய பாரம்பரிய மிக்க தமிழ்மொழியும் இவ் இணையத்தில் தனக்கென ஓர் இடத்தைப் பெற்று வளர்ந்து வருகின்றது. இணையத்தில் எண்ணிலடங்கா இலக்கிய வகைகளைப்பெற்று வளர்ந்து வரும் தமிழ் மொழிக்கு 'வலைப்பூக்கள்' என்ற புதிய இலக்கிய வகை தோன்றி பெரும் பங்காற்றிவருகிறது. வலைப்பூக்கள் என்றால் என்ன? அதன் தோற்றம், தமிழில் தோன்றிய வரலாறு, மற்றும் அதன் வகைகளாக இலக்கியம் சார்ந்த வலைப்பூக்கள், பக்தி, ஆன்மீகம், கணிப்பொறி, மருத்துவம், பல்சுவைச் சார்ந்த வலைப்பூக்கள் என பகுக்கப்பட்டு ஒவ்வொன்றின் தமிழ் பயன்பாட்டையும் எடுத்து விளக்க இக்கட்டுரை விளைகிறது.

வலைப்பூ

ஒரு சமுதாயம் இன்றைய பணிகளை இன்றைய கருவி கொண்டு செய்ய வேண்டும். இன்றைய பணியை நேற்றைய கருவிகொண்டு செய்யும் இனத்தின் நாளை வாழ்வு நலியும். இது தவிர்க்க முடியாதது” என்று டாக்டர் வா.செ.குழைந்தைசாமி அவர்களின் கூற்றின்படி நாம் இன்றையப் பணியை இன்றைய கருவிகொண்டு செய்ய வேண்டும். அதன் அடிப்படையில்தான் நாம் இணையத்தைப் பயன்படுத்தத் தொடங்கிவிட்டோம். அதில் வலைப்பூக்கள் என்ற ஒரு தனி இலக்கிய வகைத்தோன்றியுள்ளன. ஒருவரிடமிருந்து பிறருக்குத் தெரிவிக்கப் பயன்படுத்தப்படும் தகவல் தொடர்புக்கான எழுத்துக்கள், ஒலி, ஒளி வடிவக் கோப்புகள், ஓவியம், படங்கள் என்று அனைத்தையும் இணையம் வழியே தனிப்பட்ட ஒருவர் உலகில் இருக்கும் பிறருக்குத் தெரிவிக்க உதவும் இணைய வழியிலான ஒரு சேவையே வலைப்பூ என்பதாகும்.

வலைப்பூ என்பதை ஆங்கிலத்தில் பிளாக் (Blog) என்கிறார்கள். இதன் மூலம் வெப்பிளாக் (Webblog) என்பதாகும். 17-12-1997-ல் ஜார்ன் பெர்கர் (John Barger) என்பவர்தான் வலைப்பூவிற்கு ஆங்கிலத்தில் Webblog என்ற பெயரை உருவாக்கிப் பயன்படுத்தினார். இதன் பின்பு இதன் சுருக்க வடிவமான Blog எனும் பெயரை பீட்டர் மெர்ஹால்ஸ் (Peter Merholz) என்பவர் 1999 ஆம் ஆண்டு ஏப்ரல் மாதம் முதல் பயன்படுத்தத் தொடங்கினார். இவரது வலைப்பதிவின் பக்கப் பட்டையில் Webblog எனும் சொல் இரண்டாக உடைக்கப்பட்டு We blog என்று பிரித்துக் கையாளத் தொடங்கினார். இப்படியே வலைப்பூவிற்கு Blog எனும் பெயர் நிலைத்து விட்டது.

தமிழ் வலைப்பூ

இந்த blog எனும் ஆங்கிலச் சொல்லிற்கு இணையாகத் தமிழில் ஒரு பெயர் உருவாக்க விரும்பிய போது தமிழ் உலகம் மற்றும் ராயர் காபி கிளப் மடலாடற் குழு (மின் குழுமம்) உறுப்பினர்கள் தங்கள் கலந்துரையாடல்களின் வழியாக blog-க்கு வலைப்பூ என்று தமிழில் பெயர் உருவாக்கினர். இன்று தமிழில் இந்த வலைப்பூ என்ற பெயரே பயன்பாட்டில் இருந்து வருகிறது.

வலைப்பூ சேவை

வலைப்பூ வசதிக்கான சேவையை முதன்முதலாக 1996 ஆம் ஆண்டில் எக்ஸான்யா (Xanya) எனும் நிறுவனம் வழங்கத் தொடங்கியது. 1997 ஆம் ஆண்டில் சுமார் 100 நாட்குறிப்பேடுகள் இடம் பெற்றன. அதன் பிறகு சில நிறுவனங்கள் வலைப்பூவிற்கான இலவச இடவசதியைச் செய்து கொடுத்தன. இந்நிறுவனங்களில் ஒன்று பிளாக்கர்ஸ்.காம் எனும் பெயரில் வலைப்பூ அமைப்பதற்கான சேவையை இலவசமாக அளித்து அதிகமான வலைப்பூக்களை அமைக்க வாய்ப்பளித்தது. இதன் மூலம் ஆங்கிலத்தில் பலர் தங்களுக்கான வலைப்பூக்களை உருவாக்கத் தொடங்கினர். இதன் வளர்ச்சியைக் கண்ட கூக்ளி (Google) நிறுவனம் இந்நிறுவனத்தை விலைக்குப் பெற்றது. அதன் பிறகு அனைத்து மொழிகளிலும் வலைப்பூ அமைப்பதற்கான சேவை அளிக்கப்பட்டது.

முதல் தமிழ் வலைப்பூ

தமிழ் மொழியிலான முதல் வலைப்பூவை நவன் என்கிற வலைப்பதிவர் 2003 ஆம் ஆண்டில் ஜனவரி 26-ல் உருவாக்கினார் என்று அவருடைய வலைப்பக்கத்தில் தெரிவிக்கப்பட்டுள்ளது. ஆனால் 2003 ஆம் ஆண்டில் ஜனவரி முதல் தேதியன்றே கார்த்திக் ராமாஸ் என்பவர் முதல் வலைப்பூவை உருவாக்கினார் என்று சிந்தா நதி எனும் இணைய இதழில் சுட்டிக் காட்டப்பட்டுள்ளது. இந்த இரு வலைப்பூக்களில் நவன் வலைப்பூ பிளாக்கர்ஸ்.காம் தளத்திலும், (www.navan.name/blog/?p=18) கார்த்திக் ராமாஸ் வலைப்பூ பிளாக்கர்ஸ் எனும் தளத்திலும் பதிவு செய்யப்பட்டுள்ளன.

கார்த்திகேயன் ராமசாமி (கார்த்திக் ராமாஸ்) எனும் வலைப்பதிவர் தமிழில் செய்த முதல் வலைப்பூ என்று பேராசிரியர் மு. இளங்கோவன் எட்டாவது தமிழ் இணைய மாநாட்டு மலரில் குறிப்பிட்டுள்ளார். தமிழ் விக்கிப்பீடியாவிலும் கார்த்திகேயன் ராமசாமி வலைப்பூதான் முதல் தமிழ் வலைப்பூ என்று சுட்டிக்காட்டப்பட்டுள்ளது. (karthikramas.blogdrive.com/archive/21.html)

தமிழ் வலைப்பூக்கள் வளர்ச்சி

தமிழ் வலைப்பூக்கள் உருவாக்கம் மற்றும் பயன்கள் குறித்த கட்டுரை ஒன்று திசைகள் எனும் இணைய இதழில் வெளியானதைத் தொடர்ந்து தமிழ் வலைப்பூக்கள் குறித்துப் பலருக்கும் தெரியத் தொடங்கியது. தமிழ் வலைப்பூக்களின் தொடக்கக் காலத்தில் தமிழ் எழுத்துருப் பிரச்சனைகள் இருந்ததால் இதன் வளர்ச்சி சற்றுக் குறைவாகவே இருந்தது. 2003 ஆம் ஆண்டிலிருந்து 2005 ஆம் ஆண்டு வரை சுமார் 1000 வலைப்பூக்கள் வரையே தோன்றியிருந்தன. அதற்கடுத்து 2005 முதல் 2007 ஆம் ஆண்டு வரையான காலத்தில் இந்த எண்ணிக்கை 4000 ஆக அதிகரித்தது என்று பேராசிரியர் க.துரையாசன் எழுதிய இணையமும் இனிய தமிழும்” என்ற நூலில் குறிப்பிட்டிருக்கிறார். அதற்கடுத்து தமிழ் வலைப்பூக்களின் எண்ணிக்கை வேகமாக அதிகரித்து 12000-ஐத் தாண்டிவிட்டது. இது இன்னும் பன்மடங்காக உயரக் கூடும்.

தமிழ் வலைப்பூக்களின் வகைப்பாடுகள் வளர்ச்சியும்

தமிழ் வலைப்பூக்களின் உள்ளடக்கத்தைக் கொண்டு முக்கியமான சில தலைப்புகளின் கீழ் அவற்றை வகைப்படுத்தலாம்.

1. தமிழ் வலைப்பூக்களில் அதிகமாக கவிதைகளுக்கான வலைப்பூக்கள் இருக்கின்றன. வலைப்பூக்களை உருவாகியிருக்கும் பல வலைப்பதிவர்கள் தங்கள் கவிதைகளை அவர்களுக்கான வலைப்பூக்களில் அதிக அளவில் வலையேற்றம் செய்து வருகிறார்கள்.

எடுத்துக்காட்டாக கவினுலகம் என்னும் வலைப்பூ முனைவர் நா. கண்ணன் அவர்களால் உருவாக்கப் பட்டது. இந்த வலைப்பூவில் இலக்கியம் சார்ந்த பல்வேறு கருத்துக்கள் கட்டுரை வடிவிலும், கருத்துரையின் மூலமும் பதிவேற்றம் செய்யப்பட்டு வருகிறது. ஜூலை 2003-ஆம் ஆண்டு உருவாக்கப் பட்டுள்ளது. (www.emadal.blogspot.com)

2. இந்தக் கவிதைகளுக்கான வலைப்பூக்களைத் தவிர தமிழாசிரியர்களாகவும், பேராசிரியர்களாகவும் பணியாற்றி வரும் சிலர் தமிழ் இலக்கியம் சார்ந்த கருத்துக்களைப் பதிவேற்றி வருகின்றனர். மானிடன் என்ற பெயரில் தமிழ் இணைப் பேராசிரியருமான முனைவர் மு. பழநியப்பன் அவர்களால் இவ்வலைப்பூ 25.04.06-ல் உருவாக்கப்பட்டது. இவ்வலைப்பூவில் அதிக அளவில் கவிதைகளும் கட்டுரைகளும் இருக்கின்றன. தமிழ் இலக்கியங்களின் பரிமாணங்களைத் தனக்கே உரிய புதிய சிந்தனைகளுடன் இங்கு பதிவேற்றம் செய்துள்ளார். (www.manidar.blogspot.com) மு இளங்கோவன் என்ற பெயரில் ஒரு வலைப்பூ 1.5.2007 முதல் இயங்கி வருகிறது. பேராசிரியர் மு. இளங்கோவனால் வெளியிடப்படுகிறது. இதில் 300 இடுகைகள் வரை இடம் பெற்றுள்ளன. இவர் நாள்தோறும் புதிய புதிய இடுகைகளைப் பதிவேற்றம் செய்த வண்ணம் உள்ளார். இவரது கட்டுரைகள் இலக்கியத்தரம் வாய்ந்ததும் தெளிந்த நடையுடையதுமாக அமைந்துள்ளன. இவ்வலைப்பூவிலிருந்து பிற வலைதளங்களுக்குச் செல்லும் இணைப்பு வசதியும் செய்யப்பட்டுள்ளது. இங்கு பழம்பெரும் இலக்கியவாதிகளின் தொகுப்புகள் தொகுத்தளிக்கப்பட்டிருக்கிறது. (mwww.mvelangovan.blogspot.com)

3. ஆன்மீக ஈடுபாடுடைய பலர் அவரவர்க்குப் பிடித்த ஆன்மீகக் கருத்துக்களை வலியுறுத்தும் விதமாக இந்து, இசுலாம், கிறித்தவம், பவுத்தம் மற்றும் பிற ஆன்மீகக் கருத்துக்களைக் கொண்டு தமிழில் வலைப்பதிவு செய்து வருகின்றனர். கந்தர் அலங்காரம் என்ற பெயரில் கண்ணதாசன் மற்றும் ரவிசங்கர் என்பவர்களால் 2005-ல் தொங்கப்பட்ட இவ்வலைப்பூ உலகத் தமிழர்களால் வெகுவாகப் பாராட்டப்பட்டதாகும். இந்த வலைப்பூ தவிர இந்து மதத்தின் மேல் கொண்ட ஆர்வத்தின் காரணமாக திருப்பள்ளியெழுச்சி என்று மற்றொரு வலைப்பூவையும் இவர் உருவாக்கியுள்ளார். முருகப் பெருமானின் புகைப்படங்கள், பெருமைகள், விஷ்ணு பகவான் குறித்த பல செய்திகளும் சுப்ரபாதம், தோத்திரங்கள் என்று பக்தியின் உயர்வு நிலையைத் தாங்கி வெளிவந்து கொண்டிருக்கிறது. (www.murugaperuman.blogspot.com)

4. இணையப் பயன்பாட்டில் அதிகமாகப் பங்கு கொள்ளும் கணினிக்கான தொழில்நுட்பப் பணியிலிருக்கும் பலர் கணினி தொழில்நுட்பங்களைப் பகிர்ந்து கொள்ளும் விதமாக உருவாக்கிய பல தமிழ் வலைப்பூக்கள் இருக்கின்றன. அவைகளில் சில மென்பொருட்கள் ஏப்ரல் 2005 முதல் தொடங்கப்பட்ட இவ்வலைப்பூவில் தமிழில் கணிப்பொறியை எவ்வாறு இயக்குவது, தமிழ் மென்பொருட்களின் பட்டியல்கள், என பல தகவல்கள் அடங்கிய கட்டுரைகள் உள்ளன. கணினி, இணையம் பற்றிய சில செய்திகளும் இந்த வலைப்பூவில் தரப்பட்டுள்ளன. (www.tamiltools.blogspot.com)

5. விண்வெளி, அறிவியல், கணிதம் மற்றும் நவீன தொழில்நுட்பங்களை வெளிப்படுத்தும் சில வலைப்பூக்களையும் தமிழில் சிலர் உருவாக்கியுள்ளனர்.

குருவிகள் என்ற பெயரில் ஜூலை 2003லிருந்து தொடங்கப்பட்ட இவ்வலைப்பூவில் அறிவியல் செய்திகள் குறித்த கட்டுரைகளும், புகைப்படங்களும் அதிகமாக இடம் பெற்றுள்ளன. இந்த

வலைப்பதிவர் குறித்த தகவல்களை அறிய முடியவில்லை. தமிழ் சினிமா செய்திகள் அவை குறித்த விமர்சனங்கள், நகைச்சுவை, அறிய புகைப்படங்கள் எனும் பார்வையிலான செய்திகளுடன் பல தமிழ் வலைப்பூக்கள் இருக்கின்றன.

மருத்துவக் குறிப்புகள், மருந்துகள் அதைப் பயன்படுத்தும் முறைகள் என்று மருத்துவம் சார்ந்த சில வலைப்பூக்கள் தமிழில் உள்ளன. இந்த தமிழ் வலைப்பூக்களில் சித்த மருத்துவம், ஆயுர்வேதம், ஓமியோபதி மற்றும் இயற்கை மருத்துவங்களிலான வலைப்பூக்களே இருக்கின்றன. மூலிகை வளம் என்ற பெயரில் குப்புசாமி என்பவரால் 2007-ல் உருவாக்கப்பட்டுள்ள இந்த வலைப்பூவில் பல மூலிகைச் செடிகள் குறித்தும் அவற்றின் தாவரப் பெயர், தாவரக் குடும்பம், வழக்கத்திலிருக்கும் அதற்கான வேறு பெயர்கள், பயன் தரும் பாகங்கள் போன்றவைகளைப் புகைப்படத்துடன் தந்துள்ளார். இவை தவிர பல நோய்களுக்கு மூலிகை மருந்துகள் குறித்த தகவல்களும் கொடுத்திருப்பது நல்ல பயனுள்ளதாக இருக்கிறது. (www.mooligaivazam-kuppusamy.blogspot.com)

6. பெண் உடல் நலம், பெண்களுக்கான சுதந்திரம், வேலைவாய்ப்பு போன்ற ஒரு சில பெண்களுக்கான சிறந்த வலைப்பூக்களும் தமிழில் உருவாகியிருக்கின்றன.

சாதனைப் பெண்கள் என்ற பெயரில் ஜெர்மனியிலிருக்கும் திருமதி. சந்திரவதனா செல்வகுமாரால் உருவாக்கப்பட்ட சில வலைப்பூக்களில் இதுவும் ஒன்று. இந்த வலைப்பூவில் பல அச்சிதழ்களில் வெளியான சில முக்கியப் பெண்மணிகளைப் பற்றிய செய்திகளைத் தொகுத்தளித்துள்ளார்.. (www.vippenn.blogspot.com)

தமிழ் வளர்ச்சியில் வலைப்பூக்கள்

- I. வலைப்பூக்களின் வருகையால் தமிழ் மொழி இலக்கியங்கள் வெளியுலக மக்களுக்குத் தெரிய வருகின்றன.
- II. தமிழில் இணையத்தில் எழுதுபவர்கள் பெறுகியுள்ளனர். இதனால் தமிழின் வளர்ச்சி உயர்ந்துள்ளது.
- III. வலைப்பூக்களால் உலக நாடுகள் பலவற்றில் வாழும் தமிழ் மக்களின் கருத்துக்கள் மிக விரைவாக கிடைக்கின்றன.
- IV. இலங்கை,மலேசியா, கனடா,தென்கொரியா, சிங்கப்பூர், அரபு நாடுகள் போன்றவற்றில் வாழும் மக்களின் படைப்புகள் தமிழ்மொழியில் இருப்பதால் அனைவரும் கருத்தைப் பகிர்த்துகொள்ள முடிகிறது.
- V. தமிழ் மொழியின் இலக்கண, இலக்கியங்களானா சங்க இலக்கியம் முதல்கொண்டு இக்கால இலக்கியங்கள் வரை வலைப்பூவினால் உலகத் தமிழர்களுக்குக் கிடைக்கிறது. இதனால் தமிழ் மொழி வளர்ச்சிப் பெற்றுவருகிறது.
- VI. இவைகள் அன்றி கணிப்பொறிச் சார்ந்த தகவல்கள் அதிகம் கிடைக்கின்றன. அறிவியல், விஞ்ஞானக் கருத்துக்களும் அது தொடர்பான புதிய கண்டுபிடிப்புகளும் நமக்குக் கிடைக்கின்றது.
- VII. உலக நாடுகளில் உள்ள சைவ மடாலயங்களும், திருத்தலங்களும் பற்றியச் செய்திகள் இடம்பெற்றுள்ளன.
- VIII. தமிழ் ஆய்வுக்கட்டுரைகள் அதிகம் வலைப்பூக்களில் வெளிவருகின்றன. இதனால் தமிழ் ஆய்விற்குப் பல வழிகளில் பயன்படுகின்றன.

- IX. வலைப்பூவினால் தொழில் நுட்ப வளர்ச்சிச் பெற்றுத் தமிழ்மொழி வளர்ந்து வருகின்றது.
- X. வலைப்பூகளில் வெளிவரும் படைப்புகளும், கட்டுரைகளும், கவிதைகளும், பிற கருத்துக்களுக்கும் உடனுக்குடன் பின்னூட்டம் என்ற பெயரில் விமர்சனங்கள் பல நாடுகளிலிருந்து எழுதுகின்றனர். இது தமிழ் மொழிக்குக் கிடைத்த விமர்சன இலக்கியம் என்றே கூறலாம். மேலும் பல துறைகளைச் சார்ந்த அறிஞர் பெருமக்களும் தமிழ் மொழிக்குத் தன்னால் இயன்ற பணிகளையும் செய்து வருகின்றனர்.

முடிவுரை

எப்படி இலக்கிய வரலாற்றில் சங்க காலம், சங்கம் மருவிய காலம், பக்தி இலக்கிய காலம், காப்பியக் காலம், சிற்றிலக்கிய காலம், ஐரோப்பியர் காலம் என்கின்றோமோ அதனைப் போன்று இன்றைய காலக் கட்டத்தைக் “கணினியுக் காலம்” அல்லது “தமிழ் இணையக் காலம்” எனலாம். புதிய இலக்கிய வகையாக வலைப்பூ உருவாகி உலக மொழிகளில் தமிழின் பெருமையை நிலைநாட்டிக் கொண்டிருக்கிறது. இதனால் பலவகைப்பட்ட தமிழ் இலக்கியங்கள் வெளி உலகுக்கு விரைவாகக் கொண்டுசெல்லப்படுகிறது. இதனால் தமிழ் மொழி வளர்ச்சிக்கு வலைப்பூக்களின் பங்களிப்பு அளப்பரியாத் தொண்டினைச் செய்து வருகிறது எனலாம்.

அதிகார மையங்களும் வலைப்பதிவுகளும் (Blogs)

முனைவர் நா.இளங்கோ

nagailango@gmail.com

இணைப்பேராசிரியர், பட்டமேற்படிப்பு மையம்

புதுச்சேரி-8.

தகவல் தொடர்புச் சாதனங்கள்:

தகவல் தொடர்புச் சாதனங்கள் உலகத்தை ஒரு கிராமமாகச் சுருக்கிவிட்டன. செயற்கைக் கோள்களும் இண்டர்நெட்டும் தகவல் தொடர்பு உலகத்தில் நுழைந்த பிறகு உலகத்தின் எல்லைகள் சுருங்கிக் கொண்டே வருகின்றன. சமீபத்திய வரவான செல்ஃபோன் உலகை உள்ளங்கைக்குள் சுருக்கிவிட்டது. இன்றைய சூழலில் தகவல் தொடர்புச் சாதனங்கள் இல்லாத உலகை நம்மால் கற்பனை செய்து பார்ப்பது கூட இயலாததாகி விட்டது. நாம் இப்போது தொடர்புச் சாதனங்களான ஆன ஊடகங்களுக்குள் வாழ்கிறோம். ஊடகங்கள் நமக்குத் தகவல்களைத் தருகின்றன. பொழுது போக்க உதவுகின்றன. இத்தோடு ஊடகங்கள் நிறுத்திக் கொள்வதில்லை. ஊடகங்களின் அதிகாரம் இங்கேதான் செயல்படுகின்றது. நம் வாழ்க்கையை, நம் சிந்தனையை, நம் தேவைகளைத் தீர்மானிக்கும் சக்தியாகவும் ஊடகங்களே விளங்குகின்றன.

ஊடகங்கள் உலகைப் பற்றிய தகவல்களைச் செய்தியாகவும் பிற வடிவத்திலும் தருவதோடு நிறுத்திக் கொள்வதில்லை. மாறாக உலகை எப்படிப் பார்க்க வேண்டும், எப்படிப் புரிந்துகொள்ள வேண்டும் என்பதையும் தீர்மானித்து நம்மீது அதிகாரம் செலுத்துகின்றன. உலக நிகழ்வுகளில் எவை எவை முக்கியத்துவம் உடையவை, எவை எவை முக்கியத்துவம் அற்றவை என்பதை யெல்லாம் தீர்மானிக்கும் சக்தியாக ஊடகங்கள் விளங்குகின்றன. நாம் எதைப் பற்றிப் பேச வேண்டும் எதை விட்டுவிட வேண்டும் என்பதையும் ஊடகங்களே முடிவு செய்கின்றன.

மனிதன் எழுதக் கற்றுக் கொண்டதும் எழுத்துவழித் தகவல் தொடர்பு கொள்ளத் தொடங்கியதுமான வரலாறு சில ஆயிரம் ஆண்டுகள் பழமை உடையது என்றாலும் தகவல் தொடர்பு ஊடகங்களின் பாய்ச்சல், மனிதன் காகிதத்தில் அச்சிடக் கற்றுக் கொண்டதிலிருந்தே தொடங்குகிறது.

ஊடகங்களின் அதிகாரம்:

ஊடகங்களில் வெளிப்படும் அதிகாரம் இரண்டு நிலைகளில் செயல்படும். ஒன்று, ஊடக உடைமையாளர்கள் தகவல்கள் மீதும் தகவல் தருபவர் மீதும் அதிகாரம் செலுத்துவது. மற்றொன்று, ஊடகங்களில் வெளிப்படுத்தப்படும் தகவல்கள் வாசகரிடம் அதிகாரம் செலுத்துவது. தொடக்கக் காலத்தில் தமிழகத்தைப் பொறுத்த மட்டில் அச்சு இயந்திரங்களாகிய உற்பத்திக் கருவிகளும் அச்சிட வேண்டிய தகவல்களை எழுதித் தர உதவும் கல்வியும் உயர் வர்க்கத்தினரிடம் மட்டுமே இருந்தன. எனவே தமிழகத்தைப் பொறுத்தமட்டில் இருபதாம் நூற்றாண்டின் தொடக்கத்தில் அச்சு ஊடகங்களின் வழி கருத்தியல் அதிகாரம் செலுத்துவோராக உயர் சாதியினராகவும் உயர் வர்க்கத்தினராகவும் இருந்த ஒரு சிறுபான்மைக் கூட்டத்தினரே இருந்தனர்.

ஆங்கிலேயர் தந்த கல்வியும் ஆங்கிலவழிக் கல்வியும் இருபதாம் நூற்றாண்டில் பரவலான போது எழுத்தறிவும் எழுதும் மற்றும் வாசிக்கும் பழக்கமும் அதிகமானது. ஒடுக்கப்பட்ட சமூகத்தினர் கல்விஅறிவு பெறத்தொடங்கி அச்சு ஊடகத் தகவல்களை வாசிக்கத் தொடங்கிய பிறகுதான் அச்சு ஊடகங்களின் அதிகாரம் கவனம் பெறத் தொடங்கியது. அச்சு ஊடக உரிமையாளர்கள் மற்றும் இதழாசிரியர்களின் அதிகாரம் படைப்பையும் வாசகனையும் வெகுவாகப் பாதிக்கும் தன்மை வெளிச்சத்துக்கு வந்தது. வெற்று

இலக்கியங்களும் துணுக்குத் தோரணங்களும் அதிகாரத்தை இனங்காட்டாத மலிவான ரசனைப் போக்கும் எழுத்துக்களாக்கப்பட்டன. வெகுஜனங்கள் மத்தியில் எது அதிக விலைபோகுமோ அதனையே அச்சு இயந்திரங்கள் கக்கத் தொடங்கின. அச்சு ஊடகங்கள் வணிகமயமாயின. விலைபோகும் சரக்குகள் என எழுத்துக்கள் முத்திரை குத்தப்பட்டன. தீவிரமான எழுத்துக்களும் ஒடுக்கப்பட்ட மக்களின் வாழ்க்கையும் எழுத்துக்களும் விலைபோகாச் சரக்குகள் ஆக்கப்பட்டன.

இத்தகு சூழலில்தான் சிறுபத்திரிக்கைகள் தோற்றம் பெற்றன. வணிகமயம் பின்னுக்குத் தள்ளப்பட்டு தீவிரமான எழுத்துக்களும் சோதனை முயற்சிகளும் விளிம்புநிலை மக்கள் ஆக்கங்களும் அச்சில் இடம்பிடித்தன.

சிறுபத்திரிக்கைகளில் அதிகாரம் இடம்பெயர்ந்தது. ஊடக முதலாளிகளின் இடத்தைக் குழுவும் குழுவாதங்களும் பிடித்தன. சிறுபத்திரிக்கைகள் கருத்து ரீதியான அதிகாரத்தைப் படைப்பாளிகளிடத்தும் வாசகர்களிடத்தும் செலுத்தின. அவை அறிவு ஜீவிகளின் தன்முனைப்பு மோதல் களங்களாயின. தம்மைத் தாமே புகழ்ந்து கொள்வதும் பிறரை மட்டம் தட்டுவதுமே படைப்புகளில் மேலோங்கின. கணிப்பொறி சார்ந்த D.T.P. தொழில்நுட்பத்தின் வருகை ஆளுக்கொரு இதழ், ஆளுக்கொரு குழு என்ற போக்குகளுக்குத் துணைசெய்தது. அச்சு ஊடகங்களின் அதிகாரம் தொடர்கதையானது.

புதிதாய் வருகிற படைப்பாளிகளுக்கு அவ்வளவு சுலபத்தில் ஊடகங்கள் இடமளித்து விடுவதில்லை. ஊடகங்கள் பிரபலங்களை வைத்துக் காசு பார்க்கும் வணிக நிறுவனங்களாக மாறிப் போயின. புதியவர் எழுத்துக்களின் மீது குழுவாதம், மதம், சாதி, கட்சி, இயக்கம், சித்தாந்தம் முதலான பலவும் அதிகாரம் செலுத்தும் மையங்களாயின. ஒருவர் எழுத்தின் மீது தேர்ந்தெடுத்தல், வடிகட்டல், திருத்துதல், நீக்குதல், சான்றளித்தல் என்று அதிகாரம் செலுத்த பிறர் யார்? அந்த அதிகாரத்தை அவருக்குக் கொடுத்தது யார்? அவருக்கு என்ன தகுதி? என்ற விடை தெரியாத வினாக்கள் பலப்பல. இத்தகு குழுவாதம், மதம், சாதி, கட்சி, இயக்கம், சித்தாந்தம் முதலான ஊடக அதிகார மையங்களை உடைத்தெரியும் புதிய படைப்புலக வடிவம்தான் வலைப்பதிவுகள்.

வலைப்பதிவுகள்: (Blogs)

தகவல் தொழில்நுட்பத்தின் அதி நவீன மின்னணு ஊடகமான கணினி மற்றும் செயற்கைக் கோள்கள் இவற்றின் இணைப்பால் சாத்தியமாகும் இணையம் தகவல் தொழில்நுட்ப வரலாற்றில் ஒரு புரட்சி அத்தியாயம் என்றால் அது மிகையன்று. உலகக் கணிப்பொறிகளை இணைத்துத் தகவல் பரிமாற்றம் செய்துகொள்ள உதவும் இணையம் உலகை ஒரு மேசையளவிற்குச் சுருக்கிவிட்டது. கொட்டிக் கிடக்கும் அளப்பரிய தகவல்கள், இருமுனை மற்றும் பல்முனைத் தொடர்பு, பல்ஊடகத் தொழில்நுட்பம், வேகம், உலகமொழிகளைக் கையாளும் யுனிகோட் குறிமுறை முதலான பல்வேறு சாத்தியக் கூறுகள் இணையத்தின் மிகப்பெரிய வெற்றிக்கு அடிப்படைகள். இணையம் வழங்கும் மின் அஞ்சல், இணைய அரட்டை, இணைய வணிகம், கோப்புகள் பரிமாற்றம் முதலான பல்வேறு சேவைகளில் அதிக கவனத்தைப் பெற்று உலகளாவிய வலைத்தளச் சேவை W.W.W. என்றழைக்கப்படும் (World Wide Web) சேவையாகும். வலைத்தளச் சேவையின் ஒரு பிரிவாகத் தோற்றம் பெற்று இன்றைக்குத் தனித்தொரு இணையச் சேவையாகப் புகழ் பெற்றிருப்பதுதான் வலைப்பதிவுகள் என்றழைக்கப்படும் Blogs ஆகும்.

வலைப்பதிவு என்பதற்கு இணைய அகராதி விக்கிபீடியா தரும் விளக்கம்,

“வலைப்பதிவு (Blog) என்பது அடிக்கடி இற்றைப்படுத்தப் (uptodate) படுவதற்கும், கடைசிப்பதிவு முதலில் வருமாறு ஒழுங்கு படுத்தப்படுவதற்குமென சிறப்பாக வடிவமைக்கப்பட்ட தனிப்பட்ட வலைத்தளமாகும். இற்றைப்படுத்தப் படுவதற்கும் பராமரிப்பதற்கும் வாசகர் ஊடாடுவதற்குமான வழிமுறைகள் வலைத்தளங்களைக் காட்டிலும் வலைப்பதிவுகளில் இலகுவானதாக வடிவமைக்கப் பட்டிருக்கும்.”

என்பதாகும். மேலே விக்கிபீடியா தரும் விளக்கத்திலிருந்து வலைப்பதிவின் தனித்தன்மைகளாக இரண்டு வசதிகளைச் சிறப்பித்துச் சொல்லமுடியும். 1. அடிக்கடி இற்றைப் படுத்தப்படுவது. 2. வாசகர் ஊடாடுவதற்கான வசதியினைப் பெற்றிருப்பது. இந்த இரண்டு அம்சங்கள்தான் வலைப்பதிவுகளின் தனிப்பெருஞ்சிறப்புகள்.

வலைப்பதிவு: தம் படைப்புகளைத் தாமே இணையத்தில் பதிப்பிக்கும் வசதி. ஆங்கிலத்தில் இதனை Blogging என்பர். தமிழில் இது வலைப்பதிவு எனப்படும். இதனை வலைப்பூக்கள் என்றும் சிலர் வழங்குவர். ஒருவர் தம் பெயரில் ஒரு வலைப்பதிவை உருவாக்கத் தேவைப்படுவன, கொஞ்சம் கணினி அறிவு, இணையத் தொடர்புள்ள கணினி இவை இரண்டு மட்டுமே. வலைப்பதிவுக்குப் பொருள் செலவு ஏதும் கிடையாது. இணையத்தில் இந்தச் சேவை இலவசமாகவே வழங்கப்படுகிறது. யுனிகோட் குறிமுறையைப் பயன்படுத்தித் தமிழிலேயே ஒருவர் தம்முடைய படைப்புகளைப் பதிப்பிக்கலாம். வலைப்பதிவுகளை வேறுவிதமாகவும் விளக்கலாம். அதாவது இணையத்தின் வழி ஒரு தனிநபர் உருவாக்கும் இதழ் அல்லது நாட்குறிப்பு. இந்த நாட்குறிப்பு அனைவரும் படிப்பதற்கானது.

தினமும் ஆயிரக்கணக்கானோர் தங்கள் வலைப்பதிவுகளில் பல்வேறு பதிவுகளைப் பதித்து வருகின்றார்கள். இதில் பலர் கணிப்பொறி, இணையத் தொழில் நுட்பம் அறியாதவர்கள். வலைப்பதிவாளர்களுக்குப் பயன்படும் வகையில் பல்வேறு புதிய எளிய தொழில்நுட்பங்கள் தினந்தோறும் உருவாகிக் கொண்டேயிருக்கின்றன. பெரும்பாலும் இத்தகு தொழில்நுட்ப உதவிகள் அனைவருக்கும் இணையத்தில் இலவசமாகவே கிடைக்கின்றன. கணினி பற்றிச் சிறிதளவே தெரிந்தவர்கள் கூட, தங்களுக்கென்றுச் சொந்தமான வலைப்பதிவினை உடனே உருவாக்கிக் கொள்ள முடியும்.

அநேகமாக ஒவ்வொரு வலைப்பதிவும் வாசகர்களை இலக்காகக் கொண்டே எழுதப்படுகின்றன. ஒவ்வொரு வலைப்பதிவுக்கும் தனித்தொரு வாசகர்வட்டம் அமைந்து விடுவதுண்டு. இக்காரணம் பற்றியே வலைப்பதிவுகள் வாசகர்கள் கருத்துரையாடுதற்கு ஏற்றார்போல் அமைக்கப்படுகின்றன. பதிவுகளைப் படித்த வாசகர்கள் அதற்கான தமது எதிர்வினையை, கருத்துக்களைப் பின்னூட்டங்களாக உடனடியாக அவ் வலைப்பதிவில் பதிவுசெய்து கொள்ளக்கூடிய வசதி வழங்கப்பட்டிருக்கும்.

பின்னூட்டங்களையும் அடுத்துவரும் வாசகர்கள் பார்க்க வலைப்பதிவுகளில் வாய்ப்புண்டு. தேவையேற்படும் போது பின்னூட்டம், பின்னூட்டங்களுக்குப் பின்னூட்டம் என்று சங்கிலித் தொடர்போல் பதிவு தொடர்ந்து சென்றுகொண்டே யிருக்கும்.

தகவலின் இடையிடையே படமோ, ஒலியோ, சலனப்படமோ எது தேவையோ அதனை இணைத்துத் தரும் பல்லாடகத் தகவல் முறை வலைப்பதிவுகளில் சாத்தியம். அச்ச ஊடகங்களில் எழுத்தோடு படங்களை மட்டுமே இணைக்க முடியும்.

வலைப்பதிவில் நாம் இதற்குமுன் எழுதிய அனைத்துத் தகவல்களும் தனியே வார வாரியாகவோ, மாத வாரியாகவோ வகைப்படுத்தி சேமிப்பகம் (archive) பகுதியில் பாதுகாத்து வைக்கப்பட்டிருக்கும். தேவைப்படுவோர் பழைய தகவல்களையும் இந்தப் பகுதியில் இருந்து படித்துக் கொள்ளலாம்.

வலைத்தளங்கள் - வலைப்பதிவுகள் ஒப்பீடு:

இணையத்தின் மிக முக்கிய அங்கமான வலைத்தளங்களிலிருந்து வலைப்பதிவுகள் வேறுபட்டவை. வலைத்தளங்கள் அமைத்துக்கொள்ள இடம்பிடிப்பது, வடிவமைப்பது போன்ற பணிகளுக்குக் கட்டணம் வசூலிப்பதுண்டு. ஆனால் வலைப்பதிவுச் சேவைகள் முற்றிலும் இலவசமானது. வலைத்தளங்களுக்கும் வலைப்பதிவுகளுக்கும் இடையிலான சில வேற்றுமைகளைப் பின்வரும் பட்டியல் தெளிவுபடுத்தும்.

வலைத்தளங்கள்: வலைத்தளங்களை உருவாக்க html அறிவு ஓரளவேனும் தேவை.

வலைப்பதிவுகள்: வலைப்பதிவுகளை உருவாக்க html அறிவு தேவையில்லை. வலைப்பக்கங்களை உருவாக்குவது மிகவும் எளிது. ஏற்கனவே உருவாக்கப் பட்டிருக்கும் படிவங்களில் உள்ளடக்கத்தை உருவாக்கி சமர்ப்பித்து விட்டால் தானாக வலைப்பதிவு ஒன்று உருவாக்கப்பட்டுவிடும். வார்ப்புருக்கள் (Templates) இந்தப் பணியைச் செய்து முடிக்கின்றன.

வலைத்தளங்கள்:

வலைத்தளத்திற்கான உள்ளடக்கங்களை உருவாக்கி எழுதுபவர் ஒருவராகவும், html கொண்டு அத்தகவல்களை எழுதி உள்ளிட்டு வடிவமைப்பவர் வேறு ஒருவராகவும் இருப்பர்.

வலைப்பதிவுகள்: வலைப்பதிவுக்கான உள்ளடக்கங்களை எழுதுபவரே உள்ளீடு செய்பவராகவும் இருப்பார். எந்தத் தனிப்பட்ட மென்பொருளும் தேவையில்லை. வலைப்பதிவு சேவையை வழங்குபவரே இதற்கான அனைத்து வசதிகளையும் உருவாக்கி வைத்திருப்பார்.

வலைத்தளங்கள்: வலைத்தளங்கள் அடிக்கடிப் புதுப்பிக்கப் படுவதில்லை. சில தளங்கள் மட்டுமே அத்தகைய வசதியைப் பெற்றிருக்கும்.

வலைப்பதிவுகள்: வலைப்பதிவுகள் அன்றாடம் புதுப்பிக்கப்பெறும். தேவைப்பட்டால் ஒருநாளில் பலமுறைகூட புதுப்பிக்கப்பெறும். எப்பொழுதாவது ஒருமுறை புதுப்பிக்கப்படும் பதிவுகளும் உண்டு.

வலைத்தளங்கள்: வலைத்தளங்களில் பெரும்பாலும் கருத்துப்பரிமாற்ற வசதி இருப்பதில்லை. மின்னஞ்சல் வழிப் பின்னூட்டம் சில தளங்களில் உண்டு.

வலைப்பதிவுகள்: வாசகர்கள் உடனுக்குடன் தமது கருத்துக்களை வலைப்பதிவிலேயே பதிவுசெய்யும் வசதி உண்டு. வாசகர் பின்னூட்டங்கள் ஒரு விவாதம் போலத் தொடரவும் பதிவுகளில் வாய்ப்புண்டு.

மேலே பட்டியலிடப்பட்ட வேறுபாடுகள் மட்டுமின்றி வலைப்பதிவுகளுக்கென்றே சில தனித்த வசதிகளும் இணையத்தில் உண்டு.

வலைப்பதிவுகள் -சில சிறப்பு வசதிகள்:

வலைப்பதிவுகளின் இற்றைப்படுத்தல்கள் உடனுக்குடன் செய்தியோடைகள் வழியாக அனுப்பப்படும். இவ்வசதியைப் பயன்படுத்தி வாசகர்கள் தமக்குப் பிடித்த வலைப்பதிவுகளின் செய்தியோடைகளைத் தத்தமது கணினிகளில் அதற்கான மென்பொருட்களின் உதவியுடன் இணைத்துக் கொண்டு, வலைப்பதிவுகளுக்குச் செல்லாமலேயே இற்றைப்படுத்தல்களைக் கணினியில் பெற்றுக்கொள்ளலாம். இத்தகு செய்தியோடை வசதியே வலைப்பதிவுத் திரட்டிகளும் வலைப்பதிவர் சமுதாயங்களும் இணைவதைச் சாத்தியப்படுத்தியுள்ளது.

வலைத்தளங்கள் போல் அல்லாமல் வலைப்பதிவு சேவைகளைப் பல இணையதளங்கள் இலவசமாக வழங்குகின்றன. வலைப்பதிவு சேவைகளிலேயே மிகுதியும் விரும்பப்படுவது ப்ளாக்.காம் Blogger.com சேவைதான். எளிமையான அமைப்புகளுடனும் அதேசமயம் தேவையான பல வசதிகளுடன் இந்தச் சேவை வழங்கப்படுகிறது. கூகிள் (Google.com) தேடுபொறி நிறுவனம் வழங்கும் இந்த ப்ளாக்.காம் மிகுந்த நம்பகத்தன்மை உடையது என்று தமிழ் வலைப்பதிவாளர்கள் பலராலும் பாராட்டப்படுகிறது.

ஒருவர் எழுத்தின் மீது, தேர்ந்தெடுத்தல், வடிகட்டல், திருத்துதல், நீக்குதல், சான்றளித்தல் என்று அதிகாரம் செலுத்த யாரும்ற்ற அதிகார மையங்களற்ற வலைப்பதிவுகள் ஊடக வரலாற்றில் ஒரு புரட்சி. தரப்படுத்தலுக்கும் தாமதத்திற்கும் ஆளாகாமல் ஒருவரின் எழுத்து பொது வாசிப்புக்குக் காட்சிப்படுத்தப்

படுகிறது. எல்லாத் தரப்பு வாசகர்களுக்கு முன்னாலும் எழுதப்படும் எல்லாப் படைப்புகளும் ஒரே வரிசையில் காட்சிப்படுத்தும் அதிகாரமைய உடைப்பு வலைப்பதிவுகளால் சாத்தியமாகியிருக்கிறது. தமிழ்மணம், தேன்கூடு, தமிழ்ப்பதிவுகள், தமிழ்வேளி முதலான வலைப்பதிவுத் திரட்டிகள் இப்பணியை எளிதாக்கியிருக்கின்றன.

ஊடக உடைமையாளர், தரப்படுத்துநர் போன்ற அதிகார மையங்களின் இடையீடு இல்லாமல் எழுதப்படுவன எல்லாம் ஒரு நிமிடம் கூடத் தாமதமில்லாமல் மின்னெழுத்தால் அச்சிடப்படும் வாய்ப்பு. பிரபலங்களின் ஆதிக்கங்கள் நொறுங்கி, தலைப்புகளும் உள்ளடக்கங்களுமே ஒரு படைப்பை நாம் படிக்கத் தேர்ந்தெடுப்பதற்குக் காரணங்களாகும் ஜனநாயக முறையே வலைப்பதிவுகள். எழுத்தின் தகுதி, தரம் என்ற மாயத்தோற்றங்கள் உடைந்து தகவலும் தகவலின் உடனடித்தன்மையுமே முக்கியத்துவம் பெறுகின்றன.

நிறைவாக:

1. ஒருவர் எழுத்தின் மீது, தேர்ந்தெடுத்தல், வடிகட்டல், திருத்துதல், நீக்குதல், சான்றளித்தல் என்று அதிகாரம் செலுத்த யாருமற்ற சுதந்திரம்.
2. தரப்படுத்தலுக்கும் தாமதத்திற்கும் ஆளாகாமல் ஒருவரின் எழுத்து பொது வாசிப்புக்குக் காட்சிப்படுத்தப்படல்.
3. எழுத்தின் தகுதி, தரம் என்ற மாயத்தோற்றங்கள் உடைந்து தகவலும் தகவலின் உடனடித் தன்மையுமே முக்கியத்துவம் பெறல்.
4. பல்லாடக தகவல் வழங்கும் முறை, வாசகப் பின்னூட்டங்கள், அதைத் தொடர்ந்த விவாதம் என நீளும் வாய்ப்பு.
5. படைப்பாளிகளின் பிம்பங்கள் உடைந்து வாசகன் -படைப்பாளி சமத்துவம் காணும் எழுத்து ஜனநாயகம்.

வலைப்பதிவுகளின் இத்தகு அதிகாரமைய உடைப்பு ஊடகங்களின் வரலாற்றில் ஒரு பெரிய திருப்புமுனை. கணிப்பொறி, இணையம் என்ற அறிவியல் தொழில்நுட்பம் சாதித்த புரட்சி.

Tamil Blogs – Tools, Aggregators and Beyond

Kasi Arumugam

Abstract:

Evolution of Blogs in Tamil started in 2003 and is growing steadily. This article describes technical challenges faced by Tamil Bloggers at the early phase mainly caused by encoding related issues, and how Unicode got established as a standard for Tamil web. As founder-developer of the first Tamil blog aggregator Tamilmanam, the author explains the features and issues of tools and services dedicated for the Tamil Web Content Management with particular relevance to Blogs. By recording events, the Article attempts to help future web development efforts targeted for Tamil community. A few thoughts on the current trends and future possibilities are also touched upon.

The Basics of Blogs:

A Blog is a personal website that helps an individual or a team of few people to publish their content on the internet for many others to read. The fundamental differences between a typical ‘website’ and a blog (even though slowly getting blurred these days), can be explained by the following table:

Feature	A typical ‘Website’	A Blog
Updating frequency	Less frequent - some of these not updated since launch	Very frequent, sometimes several times a day
People involved	Author→ Designer→ Developer→ Publisher/ Hoster	Blogger
Run by	Institution/Business/Government	Individuals
Real-world equivalents	Magazine, Newspaper, Brochure, Book, Directory, Album, etc.	Nothing exactly. Closest is Handbill, Letters-to-the-editor, Manuscript magazine
Participation model	Readers have no direct say	Readers’ participate through comments; bring life into the system

Evolution of Blogs as alternative media around the world

Blog is a child of technology. Unlike other forms of publications that had a classical form and a modern form, blogs cannot be thought of without the ‘wired world’ that is today’s World-Wide-Web. Blogs takes away the need for an author to be at the mercy of an editor or a publisher. Blogs have brought in equality and democracy to the countless minds that aim at the ‘authorship’. Thus a blog makes the voices of weak and less-opportune people be heard – technology’s true gift to the society.

Blogs have started playing vital roles, more visibly so in developed countries, in politics, technology and issues concerning societies. Leading media websites have exclusive pages to show popular blogs.

e.g.

- New York Times – USA (<http://www.nytimes.com/ref/topnews/blog-index.html>)
- Guardian -UK (<http://www.guardian.co.uk/tone/blog>)
- The Indian Express – India (<http://www.expressbuzzblogs.com/>).
- BBC – UK (<http://www.bbc.co.uk/blogs/>)

Technological Challenges faced by early Tamil Bloggers

It is understandable that Blog authors need to have access to web-enabled computers. But, Tamil Bloggers were challenged by few more special needs: the authoring tools & displaying technologies. They had three major needs:

1. Tamil typing tools for entering their writings into the computer, with facilities to convert from one encoding to another.
2. Technologies for ensuring that those prospective readers see Tamil text properly at their desktop without any downloads/ modifying browser configurations.
3. Tamil typing tools for readers to enter comments using no new downloads/ learning.

Technology enthusiasts and volunteer teams lent great help to Tamil Bloggers in this area that brought Tamil blogs to the forefront of many Indic languages.

Tamil typing tools:

While there were many tools available for typing Tamil in computers, the Tamil blogging community mainly operated with only few tools viz. Murasu Editor, e-Kalappai and PonguTamil writer. Murasu Systems offered free downloads of Murasu Editor. E-Kalappai was made a free download thanks to a group of donors who had paid on behalf of the community. PonguTamil was an online service from suratha.com. Of these, due to the simplicity and compatibility with multiple applications, e-Kalappai was the most popular among Bloggers, particularly those who entered Tamil computing after the blog era started. PonguTamil (<http://www.suratha.com/reader.htm>) suited those without having rights to install anything on their computers.

Display – related issues:

Then came the next challenge. Hitherto, most websites displaying Tamil content had their text created with a variety of encodings. There were proprietary fonts every viewer needs to download and install for each website one visits. Even after this, there were browser settings one had to do in order to see Tamil properly rendered on their screen.

Unicode standard was already out and few websites like ezhilnila.com, thisaigal.com, etc. have been displaying Unicode Tamil content. This demonstrated that Unicode provided the answer to most of the issues concerning displaying text at reader's computer out-of-the-box – no font downloads, no tricky browser settings. This sealed the growth of other encodings for Tamil Blogs. Even few blogs operating with TSCII text had to eventually convert to Unicode.

With the growth of blogs in Tamil, Unicode became the de-facto encoding standard for the Tamil web. Today the most visited media sites such as Dinamani, Dinamalar, Dinakaran, Kumudam, Vikatan etc., in Tamil are in Unicode.

Windows 98-specific issues:

Still there was a need to support the computers running on Windows 98, where Unicode was not working straight away. Users still needed to download Unicode fonts; People offered how-to help pages for enabling Indic language setting, updating the uniscribe rendering library (USP10.dll). But still it needed a simpler solution. The web embedded font technology (WEFT) of Microsoft (simply called 'dynamic font') filled in this gap and Tamil Bloggers readily grabbed it. Of course it required certain paid technology that worked only for a given domain. Thanks to Athirampattinam Umar Thambi's domain-independent dynamic font file Thenee.eot, Tamil blogs exploited this technology to the fullest extent and all their blogs were fully working in Windows 98. Thenee was such useful tool that it supported both TSCII and Unicode in a single font file, helping many comment-writers who were using TSCII as part of their yahoo group activity.

Tamil typing tools for comment-writers

The last Tamil-specific challenge was to offer means for the blog readers to type-in their comments. Obviously it was utterly impractical to expect a casual reader to learn and install tools/settings for typing in Tamil. Being online tool running on JavaScript that every browser supported, PonguTamil JavaScript code was embedded into many comment boxes enabling Romanized input of Tamil by readers. Gopi improved this further and offered with many options at <http://www.higopi.com/ucedit/Reuse.html>

Volunteering and Community Efforts

Tamil Blogs started appearing in the scene from early 2003. There were several online resources mainly by early volunteers like Suratha Yazhvanan, Mathy Kandasamy, Umar Thambi, etc. explaining blogging in general and Tamil-specific issues in particular. These offered help and guidance for many early Bloggers.

The 'Tamil Bloggers List' at <http://tamilblogs.blogspot.com/> originally compiled by Mathy Kandasamy was the launch pad for many new Bloggers. Community blog Valaippoo (<http://valaippoo.blogspot.com/>) was a a meeting point for discussing blogging related issues while its weekly authors tried to show case interesting posts from Tamil blogs.

An article series by the author himself on essentials of blogging (தமிழில் எழுதலாம் வாருங்கள், வலையில் பரப்பலாம் பாருங்கள்) was originally published in e-sangamam e-zine and is still available at his personal blog at http://kasilingam.com/wiki/doku.php?id=tamil_blogging

Several Bloggers themselves were writing articles on blogging and computing such as one by the author himself on encoding and Unicode: என் கோடு, உன் கோடு, யுனிகோடு தனி கோடு http://kasilingam.com/wiki/doku.php?id=tamil_unicode_for_a_blogger

By mid 2004, there were around 100 blogs written in Tamil. The blogging community needed new initiatives to help and manage growth. News feed a.k.a. rss technology became available for the popular Blogger.com service and there were efforts to network Tamil Bloggers as majority of them are from Diaspora and needed a common platform to showcase and reach out to the readers.

Arrival of Aggregators

Having created and published their blog, the authors now had to overcome the next challenge: how to make themselves visible in the crowd? Unlike the conventional magazines, blogs are written with no fixed release day or time. Blog's ability to totally break the order of time needs a technological answer too. They need to keep the interested audience to know 'What is written fresh this minute? How to reach it? Who wrote it and who are all engaged in conversations on such stuff?'

Blog aggregators help answer the above questions through intelligent use of technology. www.tamilmanam.net (formerly www.thamizmanam.com, started in Aug 2004) was the first and foremost of the new genre of websites called blog aggregators. Tamil led the whole of India in the concept of blog aggregators and have pioneering efforts in this front. Blog aggregators resemble large communities on the cyberspace serving both authors and readers. They are the main streets of the blogging village and provide a meeting point; they make this momentum to sustain. Aggregators accelerate the pace of blogging movement thus plays a vital role in the growth of web pages in languages of the people.

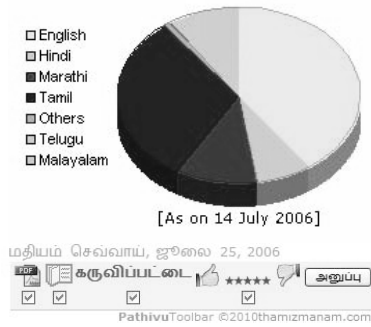
Blog portal Tamilmanam

Using open source feed aggregation technology as its starting point for Tamilmanam, the author ventured into developing programs and services tailored to meet the specific needs of Tamil bloggers. Tamilmanam employed the popular open-source tools PHP & MySQL to offer a feature-rich, contemporary service for the benefit of Tamil Bloggers.



Some of the pioneering features of Tamilmanam were:

1. Bloggers List sortable by name, location, start date, etc. updated by auto submission of URL
2. Auto aggregation of posts (updated every 20 minutes)
3. Posts written in the past days
4. Posts written on specific topics/ genre
5. Automatic aggregation of Comment status
6. Facility to convert a post into a PDF document
7. Voting on posts
8. Automatic hiding of Non-Tamil post
9. Intimation of objectionable content



Tamilmanam's most visited 'readers' page w/ star of the week, hot tags, auto-updated thumbnails of cartoons and photos - Nov-2004

Tamilmanam Star of the week showcased prominent Bloggers for a week. Tamilmanam was the first to aggregate comments to blogs long before blogging services offered newsfeeds for comments.

Tamil Blogs grew leaps and bounds with Tamilmanam. Adjoining chart published in one of the studies of blogs in India revealed the position of Tamil Blogs among other Indian language blogs.

More than a year later, Tamilmanam was rewritten and upgraded to version 2 with Pathivu toolbar. This provided for instant aggregation, post-wise categorization, one-click voting and on-the-fly pdf e-book creation was first-of-its-kind technological feature of Tamilmanam.

More Tamil Blog aggregation services

Thenkoodu was another aggregator service started around the time Tamilmanam v2 was launched but had to suspend operations due to sudden demise of the developer-owner, Sakaran. Other leading aggregator services Tamilveli and Thiratti are playing important roles in their own ways to further the growth of Tamil blogs.

Other than aggregators based on automated scripts, there were efforts by teams in showcasing select blog posts by manual recommendations. Gilli was a pioneering effort followed by Maatru . Tamilish merged the automation and manual recommendations using Digg.com model. There is a spurt of new aggregators coming out recently.

Efforts beyond aggregation

Tamilmanam's efforts towards nurturing the alternative media and supporting Tamil computing was taken further by the current team TMI (Tamil Media International Inc), a US-registered non-profit organization. TMI team also brought out Poonga, a monthly e-zine based on collections entirely out of blog posts. TMI instituted an elaborate awards program for Tamil Blogs and have developed dedicated software for managing the award program that is unique in several ways, like there is no judges panel, etc.

Tamilmanam Today

Tamilmanam currently aggregates over 7000 blogs that update with some 300+ posts everyday. It is feature-rich service and empowers a vibrant community of Tamil Bloggers. Many TMI members are part of INFITT and are involved in advancement of Tamil in internet and computing. Thanks to its primary technical contributor Sasikumar's continuous efforts, Tamilmanam is up-to-date in terms of technology. It is one of the few successful team efforts in Tamil computing world.

Tamil Blogs as alternative media

The emergence of blogs has created an independent alternative media that touch upon issues hitherto not given adequate space. The speed, economy, varieties of topics and plural perspectives possible to be written in blogs have no parallels in the conventional media. Along with blogs, the language computing is coming of age and blogs are here to stay.

Many mainstream writers have started to blog. People from other creative fields, like the film world are also attracted towards this interesting and creative pastime. Many magazines have started highlighting interesting blogs. Some of the noted blogs have been printed into books too. All this lead to greater visibility for Tamil blogs that are showing signs of alternative media.

The Trends of Tamil Bloggers:

The micro-blogging service Twitter is an effective short communication-cum-networking medium that many Bloggers are now tweeters too. The companionship they have generated through their blogs helps them get a ready-made following in Twitter. Twitter is available in cell phones, a major push over other web-only networking services. Still, for expressing elaborate thoughts on the net, Tweeters use blogs. There are several automated cross-service applications that lets one makes best use of each from the other. Many Bloggers trans-locate their friendship network into Social Networking such as FaceBook, Orkut, MySpace, etc. and Professional Networking such as LinkedIn. As in micro-blogging, their Blogging continues, albeit less frequently.

Bloggers arrange meets as an extension of their coming together in Blogosphere. Such meetings of Tamil Bloggers are very common in Chennai, as here is the largest congregation of Tamil Bloggers. Meets also happen at locations such as Coimbatore, Puducherry, Bangalore, Erode, Madurai. Bloggers from the Diaspora too meet at Singapore, USA, UAE, Sri Lanka, etc. Bloggers organize workshops to encourage and help more would-be bloggers.

Future scenario

There are blog-driven news/media sites in English, a very popular one being SlashDot (www.slashdot.com) Tamil blogs have full potential to evolve into such a format which will differ from plain aggregators/bookmarking applications. A possible blog-driven media site with following working arrangement is very much a reality:

1. Enlisted authors write independently and publish their articles at a central site on topical matters on a continuous basis.
2. Editor's panel is just to monitor and intervene on legal issues or complaints but no formal pre-reading before publication. This ensures timely publication on news and articles, but still moderated for integrity/legality.
3. This online media will have at a given time 10's or 100s of contributors, who have been selected by the editor's panel based on their demonstrated authorship. Thus there is no scope for much noise that is common in current aggregators.
4. There will be moderation of comments, voting and abuse-reporting
5. This site will be monetized by advertisements and the revenue will be shared among the contributors after deducting the expenses, proportional to the page hits of each article.

Tamil blog world is waiting for web application developers to bring out interesting relevant applications.

வளர்ந்துவரும் மலேசியத் தமிழ் இணைய ஊடகம்

சுப.நற்குணன்

மலேசியா

suba.nargunan@gmail.com

http://thirutamil.blogspot.com

முன்னுரை

தமிழ் இணைய ஊடகம் இன்று பல பரிணாமங்களைக் கண்டு மிக வேகமாக வளர்ந்து வருகின்றது. அதற்கேற்றாற்போல, மலேசியத்தைத் தளமாகக் கொண்டு செயல்படும் தமிழ் இணைய ஊடகம் மெல்லென காலான்றி வருவதைக் காண முடிகிறது. உலகத் தமிழ்க் கணிமை, தமிழ் இணைய வளர்ச்சியில் மலேசியாவின் பங்கும் குறிப்பிட்டுச் சொல்லத்தக்க அளவில் இருக்கின்றது. எனினும், மலேசியாவில் தமிழ்க் கணினி, இணைய ஊடக வளர்ச்சிக்கான வழிகாட்டல்கள் அல்லது அதற்குரிய கல்வி வாய்ப்புகள் அரிதாகவே இருக்கின்றன. தமிழ்க் கணினி, இணையத் தொழில்நுட்பத்தைக் கற்பிக்கும் நிறுவனங்கள் அல்லது அமைப்புகள் எதுவும் இல்லை எனலாம். மேலும், மக்களிடையே குறிப்பாக, இளையோர்களிடையே தமிழ்க் கணிமை, இணையத்தை அறிந்துகொள்ளும் ஆர்வமும் அதிகம் இருக்கவில்லை. இதற்குப் பல்வேறு காரணங்கள் இருப்பதை அறியமுடிகிறது. எனவே, மலேசியாவில் தமிழ்க் கணினி, இணைய ஊடகத்தை மக்களிடையே மிகப் பரவலாகக் கொண்டு செல்வதற்குத் தகுந்த திட்டமிடல்கள் உடனடியாகத் தேவைபடுகின்றன. அவ்வாறு செய்தால் மலேசியத் தமிழ் இணையம் பெரிதாக வளர்ச்சியடையும் என்பது திண்ணம். அதற்குரிய மனித மூலதனம், ஆற்றல், வாய்ப்பு, வசதி ஆகிய அனைத்தும் மலேசியத்தில் நிரம்ப இருக்கின்றது.

மலேசியாவில் தமிழ்க் கணிமை தொழில்நுட்பம்

மலேசியாவில் தமிழ்க் கணிமை மீதான ஈடுபாடு 1980களின் தொடக்கத்தில் ஏற்பட்டது. இதற்குத் தோற்றுவாயாக இருந்து செயல்பட்ட முன்னோடிகளில் முத்து நெடுமாறன் மிகவும் குறிப்பிடத்தக்கவர். இவரோடு இரவிந்திரன், மா.அங்கையன் முதலியோரும் இத்துறையில் ஆர்வத்தோடு ஈடுபாடு காட்டிய முன்னோடிகளாவர்.

மலேசியாவில் தமிழ் மென்பொருளை உருவாக்கும் பணியில் முனைந்து ஈடுபட்டு 'முரசு' செயலியை வெளியிட்டு சாதனை செய்தவர் முத்து நெடுமாறன். முரசு செயலியின் வருகைக்குப் பின்னர் மலேசியாவில் மட்டுமின்றி, உலகம் முழுமையும் தமிழ்க் கணிமை புதிய பரிணாமத்தை நோக்கி வளர்ச்சிகண்டது எனலாம். இதே காலக்கட்டத்தில் இரவிந்திரன் என்பார் 'துணைவன்' என்ற மற்றொரு தமிழ்ச் செயலியை வடிவமைத்து வெளியிட்டார்.

பின்னர், 1992ஆம் ஆண்டில் 'நளினம்' என்னும் பெயரில் இன்னொரு புதுவகை செயலியை சிவகுருநாதன் என்பவர் உருவாக்கி அளித்தார்.

இப்படியாக, தமிழ்க் கணிமை மேம்பாட்டிற்கு மலேசியக் கணிஞர்களின் பங்களிப்பு குறிப்பிடத்தக்க அளவில் இருந்திருக்கிறது. எனினும், இக்காலக்கட்டத்தில் தமிழ் தமிழ்க் கணினி, இணையப் பயன்பாடு குறிப்பிட்டுச் சொல்லத்தக்க வகையில் வளர்ச்சியை எட்டவில்லை.

மலேசிய அச்சு ஊடகம் கணினிமயமாதல்

முரசு மென்பொருள் தொடர்ந்து மேம்படுத்தப்பட்டு வெளியிடப்பட்டதன் பயனாக, அதுவரை அச்சுக் கோர்ப்பு முறையையும் தட்டச்சு முறையையும் நம்பியிருந்த மலேசிய அச்சு ஊடகம் படிப்படியாக கணினி முறைக்கு மாறியது. 1989ஆம் ஆண்டில் 'மயில்' எனும் வார இதழ் முதன்முதலாக முரசு செயலியைப் பயன்படுத்தி வடிவமைப்புச் செய்யப்பட்டு வெளிவந்தது. அதனுடைய எழுத்துரு அமைப்பு வாசகர்களிடையே நல்ல வரவேற்பைப் பெறவே, முன்னணி வார இதழாக 'மயில்' வளர்ந்தது. பின்னர், பல வார, மாத இதழ்கள் முரசு செயலியைப் பயன்படுத்தி கணினியில் வடிவமைப்புச் செய்யப்பட்டு வெளிவந்தன. காலப்போக்கில், மலேசியாவின் முன்னணி நாளிதழ் அனைத்தும் கணினி வடிவமைப்புக்கு மாறவேண்டிய நிலைமை ஏற்பட்டது.

இவ்வாறு, முரசு செயலி மலேசிய கணினித் துறையில் பாரிய மாற்றத்தை ஏற்படுத்தியது குறிப்பிடத்தக்க ஒன்றாகும். பின்னர், 1993இல் வெளிவந்த முரசு 8 செயலி மைக்ரோசாப்டு விண்டோசில் செயல்படும் தரத்தில் அமைந்திருந்தது. இதன் பிறகு, மலேசியத் தமிழிடைமே தமிழ்க் கணினி பயன்பாடு ஓரளவுக்கு வேகமாகச் சூடுபிடிக்கத் தொடங்கியது எனலாம்.

தமிழ்ப் பள்ளிகளில் கணினி

அச்சு ஊடகத்துறையை அடுத்து, கணினியை அதிக அளவில் பயன்படுத்தும் இடமாக மலேசியத் தமிழ்ப்பள்ளிகள் விளங்கின. 1995ஆம் ஆண்டு வாக்கில், மலேசிய அரசாங்கம் நாட்டில் உள்ள அனைத்துப் பள்ளிகளுக்கும் கணினிகளை வழங்கியது. இதன் பயனாக தமிழ்ப்பள்ளிகளில் முதன்முறையாகக் கணினிகள் நுழையத் தொடங்கின. ஆனாலும், தொடக்கக் காலத்தில் இந்தக் கணினிகளில் தமிழ் மென்பொருள் எதுவும் உள்ளீடு செய்யப்படவில்லை.

ஒரு சில ஆண்டுகளுக்குப் பின்னர், 1999ஆம் ஆண்டு வாக்கில் முரசு நிறுவனமும் மலேசியக் கல்வி அமைச்சும் இணைந்து தமிழ்ப்பள்ளிகளுக்கு இலவயமாக முரசு செயலியை வழங்கின. அதோடு, முரசு செயலியைப் பயன்படுத்தி தமிழில் தட்டச்சு செய்யும் வழிமுறைகளும் கற்றுத்தரப்பட்டன. இந்த நிகழ்வுக்குப் பின்னர், மலேசியாவில் தமிழ்க் கணினிமை பயன்பாடு குறிப்பிடத்தக்க அளவில் முன்னேற்றம் கண்டது. கணினித் திரையில் முதன்முறையாகத் தமிழைக் கண்டு தமிழாசிரியர்களும் தமிழ்ப்பள்ளி மாணவர்களும் பெற்றோர்களும் அகமும் முகமும் மலர்ந்து அளவில்லா மகிழ்வு எய்திய பொற்காலமாக அந்தக் காலக்கட்டத்தைச் சொல்லலாம்.

2003ஆம் ஆண்டில், கணினித், அறிவியல் ஆகிய இரண்டு பாடங்களை ஆங்கிலத்தில் கற்பிக்கும் புதிய கல்வித் திட்டத்தை மலேசியக் கல்வி அமைச்சு அறிமுகப்படுத்தியது. இந்தக் கல்வித் திட்டமானது, மலேசிய மக்களிடையே கணினி பயன்பாட்டைத் திறுமென உயர்த்தியது. எல்லாரும் கணினியை நாடி, தேடி ஓடும் நிலைமை உண்டானது. அரசு அலுவலகங்கள், சேவை நிறுவனங்கள், வணிக நிறுவனங்கள், பள்ளிகள் என எல்லா இடங்களிலும் கணினி பயன்பாடு பெருமளவில் கூடிப்போனது.

இம்மட்டில் நின்றுவிடாமல், வீட்டுக்கு ஒரு கணினி, ஆளுக்கு ஒரு கணினி என்ற அளவுக்கு குழுகாயத்தில் எல்லா தரப்பினரும் கணினியைப் பயன்படுத்தத் தொடங்கினர். இன்றையச் சூழலில் கணினி, இணையப் பயன்பாடு என்பது மக்களின் வாழ்வியலின் ஒரு அங்கமாகவே ஆகிக்கொண்டிருக்கிறது. அதற்கு ஏற்றதுபோல மலேசியாவில் நவீன மயமான கம்பியில்லா இணையச் சேவைகள் மிக வேகமாக வளர்ந்து கொண்டிருப்பதானது கணினி, இணையப் பயன்பாட்டை மென்மேலும் ஊக்கப்படுத்துவதாக உள்ளது.

மலேசியாவில் தமிழ் இணைய மாநாடு

கடந்த 2001 ஆகத்து 26 - 28 வரை மலேசியாவில் நடைபெற்ற தமிழ் இணைய மாநாடு, இங்கு தமிழ்க் கணினி, இணையப் பயன்பாட்டில் புதிய பரிமாணங்களை ஏற்படுத்தியது என்றால் மிகையன்று. இந்த மாநாட்டின் வாயிலாகக் கணினி, இணையத்தில் தமிழுக்கு இருக்கின்ற வசதிகள், வாய்ப்புகள், சிக்கல்கள்

ஆகியவை பற்றிய விழிப்புணர்வு மலேசியத் தமிழர்களிடையே மேலோங்கியது. இம்மாநாட்டுக்குப் பின்னர், தமிழ்க் கணிமை மீது மக்களின் பார்வை மிகவும் அழுத்தமாகப் பதிந்தது. தமிழ்க் கணிமை பயனாளர்கள் எண்ணிக்கையின் அடிப்படையில் பன்மடங்கு அதிகமாகினர். இத்தனை மாறுதல்கள் ஏற்படுவதற்கு மீண்டும் முகாமையான காரணமானவராக இருந்து பாடாற்றியவர் முத்து.நெடுமாறனும் அவருடன் இணைந்து வினையாற்றிய மாநாட்டுக் குழுவினரும் தான்.

மலேசியாவில் தமிழ் இணைய ஈடுபாடு

மலேசியாவில் நடந்த தமிழ் இணைய மாநாட்டுக்குப் பின்னர், தமிழ் மக்களிடையே கணிமை பயன்பாடு கிடுகிடுவென கூடிவிட்ட அதே தருணத்தில், அதன் தொடர்பாக இருக்கும் இணைய ஈடுபாடும் வளர்ந்துகொண்டே வந்தது. 2000ஆம் ஆண்டுகளின் தொடக்கத்தில் தமிழ் இணையத்தின் பயன்பாடு நாடு முழுவதும் புகழ்பெறவும் பரவலாகவும் தொடங்கியது.

ஆனாலும், தொடக்க காலத்தில் மலேசியாவிலிருந்து செயல்படும் எந்தவொரு இணையத்தளமும் இருக்கவில்லை. அயலகத் தமிழர்கள் சிலர் நடத்திய இணையத்தளங்கள் சிலவற்றை இங்கிருக்கும் சிலர் மிகுந்த ஆர்வத்துடன் படித்து வந்தனர். அதனால் ஏற்பட்ட ஆர்வத்தின் பயனாக ஓரிரு மலேசிய எழுத்தாளர்கள் அந்த இணையத்தளங்களுக்கு தங்கள் படைப்புகளை எழுதியனுப்பியுள்ளனர். அவை அவ்விணையத் தளங்களில் வெளியிடப்பட்டிருக்கின்றன.

காலப்போக்கில், தமிழ் ஆர்வத்தையும் கொஞ்சம் கணிமை தொழில்நுட்ப அறிவையும் கொண்டவர்கள் சிலர் சொந்தமாக மலேசிய மணத்துடன் கூடிய இணையத்தளம், வலைப்பதிவு, மடற்குழு என பலவற்றையும் உருவாக்குவதில் ஆர்வத்துடன் ஈடுபட்டு வெற்றிகண்டனர். இவ்வாறு சிறிய அளவில் தொடங்கப்பட்டவை இன்று பாராட்டும்படியாக வளர்ந்துள்ளனர் என்பதை கண்கூடாகக் காணமுடிகிறது. தமிழ் இணைய வெளியில் மலேசியத் தமிழர்கள் சிலரும் தரமான பங்களிப்பைக் கொடுத்துள்ளனர்; தமிழ் இணைய உலகைச் செழிக்கச் செய்திருக்கின்றனர் – செய்தும் வருகின்றனர்.

மலேசியத் தமிழ் இணையத்தளங்கள்

மலேசியாவிலிருந்து இணையத்தில் முதன்முதலாகக் கால்பதித்த முன்னோடியாகக் குறிப்பிடத்தகுந்தவர் மருத்துவர் ஐயா.சி.ஜெயபாரதிதான். இவர் 1998ஆம் ஆண்டிலேயே அகத்தியர் மடற்குழுமம் வாயிலாக அரும்பணியைச் செய்தவர். 'ஜெய்பி' என்று இணைய உலகில் அறியப்பட்ட இவர் அகத்தியர் மடற்குழுமம் வாயிலாகத் தமிழிய இந்திய வரலாறு, மலேசிய வரலாறு, கலை, இலக்கியம், பண்பாடு, சமயம், சித்திரியல், சமூகவியல், சமூகச் சீர்திருத்தம், அறிவியல், எதிர்காலவியல் ஆகிய பல்வேறு துறைகளைப் பற்றிய ஆய்வுச் செய்திகளையும் அரிய தகவல்களையும் கொண்டையளித்தவர்.

இவருக்குப் பின்னர் மலேசியாவைச் சேர்ந்த விரல்விட்டு எண்ணக்கூடிய அளவில் சில நிறுவனங்களும் ஓரிரு தனியாட்களும் இணையத்தளம் நடத்த தொடங்கினர். அப்படி இணையத்தளம் தொடங்கியவர்களும் படிக்க ஆளில்லாத காரணத்தினால் அல்லது யாராவது படிக்கிறார்களா இல்லையா என்று அறியவியலாத காரணத்தால் தங்கள் இணையத்தளத்தை மூடிவிட்டனர்.

2005ஆம் ஆண்டுக்குப் பின்னர் தமிழ் இணையம் மலேசியா, மலேசியத் தமிழ் எழுத்துலகம், மொழி.நெட், தமிழியக்கம் போன்ற இணையத்தளங்கள் உருவாகி வலம் வந்தன. இந்தக் காலக்கட்டத்தில் தமிழ் இணையத்தளங்கள் மக்களிடையே செல்வாக்குப் பெற்ற தகவல் ஊடகங்களாகச் செயல்படும் நிலைமை உருவாகி இருக்கவில்லை.

2008ஆம் ஆண்டு பிப்பிரவரி திங்கள் 24ஆம் நாள் மலேசியாவின் 12ஆவது பொதுத் தேர்தலுக்கான வேட்புமனு தாக்கல் செய்யும் நாள். இதேநாள், மலேசிய இணையத்தள வரலாற்றில் மிகவும் முக்கியமான நாளாகவும் அமைந்தது. அன்றுதான் மலேசியாஇன்று.காம் என்ற இணையத்தளம் செயல்படத் தொடங்கியது. ஏற்கனவே 1999இல் மலேசியாகினி.காம் எனும் பெயரில் ஆங்கிலம், மலாய், சீனம் ஆகிய

மும்மொழிகளில் செயல்பட்டுவந்த இது புதிதாக மலேசியாஇன்று.காம் என்று முழுக்க முழுக்க தமிழில் செயல்படத் தொடங்கியது. நாளிதழ், வானொலி, தொலைக்காட்சி முதலான மரபு ஊடகத்தையே அதுவரை நம்பியிருந்த மலேசியத் தமிழர்களுக்கு ஒரு மாற்று ஊடகமாக வாய்த்த இணையத்தளம்தான் மலேசியாஇன்று.காம்.

நவம்பர் 25, 2007இல் மலேசியத் தமிழர்கள் அரசாங்கத்திற்கு எதிராகப் பேரளவில் ஒன்றுதிரண்டு சாலைப் போராட்டத்தில் களமிறங்கிய வரலாற்று நிகழ்வுதான் பின்னாளில் மலேசியாஇன்று.காம் இணையத்தளம் உருவாகுவதற்குப் பின்னணியாக இருந்தது. அந்த நேரத்தில் மலேசியத் தமிழர்களிடையே ஏற்பட்ட அரசியல் விழிப்புணர்வையும் எழுச்சியையும் போராட்டங்களையும் உடனுக்குடன் அறிந்துகொள்வதற்கு ஒரு மாற்று ஊடகம் மிகவும் தேவைப்பட்டது. அதனை நிறைவுசெய்யும் வகையில் மலேசியாஇன்று உதயமானபோது, அதற்குத் தமிழ் மக்களிடமிருந்து பெரும் ஆதரவு கிடைத்தது. இங்குதான், மலேசியத் தமிழர்களிடையே இணைய ஊடகம் மிகப் பரவலாகப் புகழ்பெற்றது.

தற்போது மலேசியாஇன்று இணையத்தளம் மிக அதிகமான வாசகர்களோடு வெற்றிநடை போட்டு வருகின்றது. இதனைத் தொடர்ந்து வணக்கம் மலேசியா, விடுதலை மலேசியாஇன்று முதலான இணையச் செய்தி ஊடகங்கள் உருவாகின.

ஆகக் கடைசியாக, மலேசியாவின் முன்னணி நாளிதழ்கள் இரண்டு இணையப் பதிப்பாக வெளிவரத் தொடங்கியுள்ளன. மலேசிய நண்பன், மக்கள் ஓசை ஆகிய இரு நாளிதழ்கள் மின்னியல் வடிவத்தில் வலம்வந்து, மலேசிய இணைய ஊடகத்தை செழிக்கச் செய்துகொண்டிருக்கின்றன. தவிர வல்லினம் என்ற மின்னிதழ் ஒன்றும் மலேசியாவிலிருந்து இணையத்தில் வலம் வருகிறது.

மலேசியத் தமிழ் வலைப்பதிவுகள்

மலேசியாவில் இணைய ஊடகத்தை வளர்த்தெடுத்ததில் தமிழ் வலைப்பதிவுகளுக்கு மிகப்பெரும் பங்குண்டு என்பதை யாரும் மறுக்கவியலாது. வலைப்பதிவு தொழில்நுட்பத்திலும் மலேசியத் தமிழர்கள் சிலர் கால்பதித்து உலக அளவில் அறியப்பெற்றுள்ளனர் என்பது குறிப்பிடத்தக்கது. மலேசியாவிலிருந்து முதன்முதலில் தமிழ் வலைப்பதிவைத் தொடங்கியவராக திருமதி.சுபாசினி அறியப்படுகிறார். இவர் சுபா இணையம், சுபா.இல்லம் ஆகிய பெயர்களில் 27 அத்தோபர் 2003இல் வலைப்பதிவு தொடங்கியுள்ளார். இவர் பினாங்கு மாநிலத்தைச் சேர்ந்தவர். பின்னர் செருமானியத்தில் குடியேறியவர்.

இவரை அடுத்து 8 சூன் 2004இல் சொகூர் மாநிலத்தைச் சேர்ந்த வாசுதேவன் இலட்சுமணன் என்பவர் விவேகம் எனும் வலைப்பதிவைத் தொடங்கினார். அதற்கு அடுத்த நிலையில் பேரா மாநிலத்திலிருந்து சுப.நற்குணன், திருத்தமிழ் என்னும் வலைப்பதிவை உருவாக்கினார். அதே 2005 மே திங்கள் 28ஆம் நாளில் சுப.நற்குணனின் நண்பராகிய விக்கினேசு தமிழோடு நேசம் என்ற பெயரில் ஒரு வலைப்பதிவைத் தொடங்கினார். இவர்களே மலேசியத் தமிழ் வலைப்பதிவுகளின் முன்னோடிகள் ஆவர். மேற்சொன்ன நான்கு வலைப்பதிவுகளும் இன்றளவும் செயல்பட்டுக்கொண்டிருக்கின்றன.

திருத்தமிழ், ஓலைச்சுவடி, நனவுகள், தமிழியர், வாழ்க்கைப் பயணம், தமிழ் எழுத்துச் சீர்மை முதலான வலைப்பதிவுகள் மலேசியாவில் மட்டுமின்றி உலக அளவில் வாசிக்கப்படும் வலைப்பதிவுகளாகப் புகழ்ப்பெற்றுள்ளன.. இன்றைய நிலையில் 50க்கும் மேற்பட்ட மலேசியத் தமிழ் வலைப்பதிவுகள் இணைய வெளியில் உலா வந்துகொண்டிருக்கின்றன.

மலேசியத் தமிழ் வலைப்பதிவுகள் இன்று மிகச் சிறப்பான முறையில் மக்களிடையே செல்வாக்குப் பெற்று வருகின்றன என்றால் மிகையன்று. அதற்கேற்றாற்போல, மலேசியத் தமிழ் வலைப்பதிவர்களும் தரமான செய்திகளையும் தகவல்களையும் வழங்கி வருகின்றனர்.

மலேசியத் தமிழ் வலைப்பதிவுகள் பெரும்பாலும் மொழி, இலக்கியம், மரபியல், கல்வி, சமூகம் சார்ந்த செய்திகளையே உள்ளடக்கமாகக் கொண்டுள்ளன. பொழுதுபோக்கு, கேளிக்கை தன்மைகள் நிறைந்த

வலைப்பதிவுகளும் உண்டு. நல்ல தமிழ்ப் பெயர்களைக் கொண்டிருப்பது மலேசியத் தமிழ் வலைப்பதிவுகளின் ஒரு சிறப்பாகும். அரிச்சுவடி, ஈரமான நினைவுகள், தமிழ்க் கவிதை, வளமை, தமிழ் ஆலயம், தமிழ் மருதம், அரங்கேற்றம், கயல்விழி, தமிழ்ச்சுவை, கவித்தமிழ், கவிச்சோலை, திருநெறி, தமிழ்ப் பூங்கா, கருத்து மேடை, தமிழோடு உயர்வோம் போன்றவை அவற்றுள் சிலவாகும்.

வலைப்பதிவு திரட்டிகள்

மலேசியாவைத் தளமாகக் கொண்ட இரண்டு திரட்டிகள் இணையத்தில் உள்ளன. வலைப்பூங்கா, திருமன்றில் ஆகிய இரண்டும் மலேசியத் திரட்டிகளாகும். இவற்றில் மலேசிய வலைப்பதிவுகள் அனைத்தையும் ஒருசேர ஒரே இடத்தில் காணலாம். இவ்விரு திரட்டிகள் மலேசியாவில் உருவாகும் வலைப்பதிவுகளை வாசகர்களிடம் கொண்டுபோய் சேர்க்கும் பணியைச் செய்துவருகின்றன.

மலேசியத் தமிழ்ப்பள்ளிகளின் வலைமனைகள்

மலேசியாவில் செயல்படும் பல தமிழ்ப்பள்ளிகள் தற்போது சொந்தமாக இணையத்தளம் அல்லது வலைப்பதிவைக் கொண்டிருக்கின்றன. இது மிகவும் வரவேற்கத்தக்க ஒரு வளர்ச்சியாகக் கருதப்படுகிறது. குறிப்பாக, பேரா மாநிலத்தில் உள்ள 134 தமிழ்ப்பள்ளிகளைப் பற்றிய தகவல்களையும் தரவுகளையும் தாங்கி 'பேரா மாநில தமிழ்ப்பள்ளிகளின் வலைமனை' என்ற இணையத்தளம் செயல்படுவதைச் சொல்லலாம். இருந்தபோதிலும், எல்லாத் தமிழ்ப்பள்ளிகளும் இணையத்தில் கால்பதிக்கும் நிலைமை இன்னமும் வரவில்லை. எனினும், எதிர்காலத்தில் இப்படியொரு சூழல் சாத்தியமாகலாம் என்பதை இப்போதைக்கு மறுப்பதற்கு இல்லை.

நாட்டுப் பிரதமரின் இணையத்தளத்தில் தமிழ்

மலேசியாவின் பிரதமர் தனியாக இணையத்தளம் நடத்துகிறார். ஒரே மலேசியா என்பது அவருடைய இணையத்தளத்தின் பெயர். இத்தளம் தொடங்கப்பட்டபோது மலாய், ஆங்கிலம், சீனம் ஆகிய மும்மொழிகளில் மட்டுமே செயல்பட்டது, மலேசியத் தமிழர்களின் கோரிக்கைக்குச் செவிசாய்த்து மலேசியப் பிரதமர் தம்முடைய இணையத்தளத்தில் தமிழ்ப் பகுதியையும் ஏற்படுத்தியுள்ளார் என்பது குறிப்பிடத்தக்கது.

ஒலி, ஒளி ஊடகங்களின் இணையத்தளம்

மலேசியாவின் அரசு வானொலியாகச் செயல்படும் மின்னல் பண்பலை வானொலி தனக்கென தனியாக இணையத்தளம் ஒன்றைக் கொண்டிருக்கிறது. அதுபோலவே, அசுட்ரோ நிறுவனத்தின் வானவில் தொலைக்காட்சி அலைவரிசை 'விழுதுகள்' எனும் இணையத்தளத்தைப் பெற்றுள்ளது. இவ்விரு இணையத்தளங்களும் தமிழில் செயல்படுகின்றன என்பது குறிப்பிடத்தக்கது. அதுமட்டுமல்லாமல், மின்னல் பண்பலை வானொலியும் தி.எச்.ஆர் இராகா தனியார் வானொலியும் இணைய ஒலிபரப்பையும் மேற்கொண்டு வருகின்றன. மலேசியச் சூழலில், இன்றைய இணையத் தொழில்நுட்பத்தில் தமிழ்மொழி அடைந்திருக்கும் வளர்ச்சிக்கு இதுவொரு நற்சான்றாகும்.

மடற்குழுமங்கள்

மலேசிய மணத்துடன் இரண்டு மடற்குழுக்கள் இணையத்தில் இயங்குகின்றன. மலேசியத் தமிழ் வலைப்பதிவர்கள் என்பது வலைப்பதிவர்களும் அதன் வாசகர்களும் பங்கேற்கும் மடற்குழுவாகும். மற்றொன்று 'தமிழ் ஆசிரியம் மடற்குழு'. இது மலேசியத் தமிழ் ஆசிரியர்களுக்காக உருவாக்கப்பட்டுள்ள மடற்குழுவாகும்.

இப்படியாக, கடந்த 10 ஆண்டுகளில் மலேசியத் தமிழ் இணைய ஊடகம் பெரிதான வளர்ச்சியைக் கண்டிருக்கிறது. தமிழ் மென்பொருள்கள், இணையத் தளங்கள், வலைப்பதிவுகள், மின்னிதழ்கள், மடற்குழு, எனப் பல்வேறு பரிணாமங்களில் மலேசியத் தமிழ் இணைய ஊடகம் கோலோச்சி வருகின்றது. மலேசியத் தமிழர்களிடையே தமிழ் இணைய ஈடுபாடும் பங்கேற்பும் நாளுக்கு நாள் அதிகரித்து வருவதைக் காண முடிகிறது.

இருப்பினும், மலேசியாவில் தமிழ்க் கணினி பயன்பாடு என்பது அனைத்து தமிழர்களையும் அளவியதாக இருக்கவில்லை. மாறாக ஒரு குறிப்பிட்ட தரப்பினரே தமிழ்க் கணினி, இணையத்தைப் பயன்படுத்தியும் பயன்கொண்டும் வருகின்றனர். இங்கு பெரும்பாலும் கணினித் தொழில்நுட்பம் அறிந்தவர்கள் தமிழ்க் கணினி, இணையப் பயன்பாட்டில் ஈடுபடுகின்றனர் என்று சொல்வதைவிட, தமிழ் அறிந்தவர்களே இத்துறையில் ஈடுபாடு காட்டுகின்றனர் எனலாம்.

முடிவுரை

மலேசியாவில் தமிழ்க் கணினி, இணைய ஊடக வளர்ச்சிக்கான வழிகாட்டல்கள் அல்லது அதற்குரிய கல்வி வாய்ப்புகள் ஆகியன அரிதாகவே இருக்கின்றன. குறிப்பாக, தமிழ்க் கணினி, இணையத் தொழில்நுட்பத்தைக் கற்பிக்கும் நிறுவனங்கள் அல்லது அமைப்புகள் எதுவும் இல்லை எனலாம். ஆங்காங்கே, சிறுசிறு குழுக்களாகவும் அல்லது தனியாட்களும் தன்னார்வத்தின் அடிப்படையில் கணினி, இணையம் பற்றிய கல்வியையும் விழிப்புணர்வையும் வழங்கி வருகின்றனர்.

மேலும், மக்களிடையே அதுவும், இணையோர்களிடையே தமிழ்க் கணினி, இணையத்தை அறிந்துகொள்ளும் ஆர்வமும் அதிகம் இருப்பதில்லை. இதற்கு அடிப்படை காரணங்களாக, தமிழ்க் கல்வி இல்லாமை, தமிழ்க் கணினி தெரிந்தோருக்கு வேலை வாய்ப்பின்மை, பொருளியல் மதிப்பு இல்லாமை முதலான பலவற்றைச் சொல்லலாம்.

எனினும், மலேசியாவில் தமிழ்க் கணினி, இணைய ஊடகத்தை வளர்த்தெடுக்கும் முயற்சிகள் அவ்வப்போது நடந்துகொண்டுதான் இருக்கின்றன. குறிப்பாக, மலேசியத் தமிழ் வலைப்பதிவர்களின் முயற்சியில் ஆங்காங்கே தமிழ்க் கணினி, இணையப் பயிலரங்குகள் நடத்தப்பட்டு வருகின்றன.

தமிழ் இணைய ஊடகம் தொடர்பான விழிப்புணர்வு மக்களிடையே அதிகரித்துக்கொண்டிருக்கிறது. தமிழ்க் கணினிப் பயன்பாடும் மலேசியத் தமிழர்களிடையே விரிவடைந்து வருகிறது. என்றாலும், இவற்றின் வளர்ச்சியானது மிகவும் மந்தமாகவே நிகழ்கின்றது. இதனை உடனடியாக விவேகமான முறையில் விரைவுபடுத்தும் வழிவகைகளை ஆராய்ந்து தக்கவற்றைச் செய்தாக வேண்டும்.

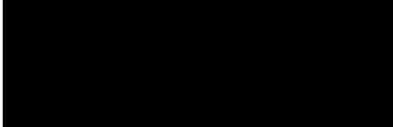
மலேசியாவில் தமிழ்க் கணினி, இணைய ஊடகத்தைத் தமிழ் மக்களிடையே மிகப் பரவலாகக் கொண்டு செல்வதற்கு தகுந்த திட்டமிடல்கள் தேவைபடுகின்றன. அவ்வாறு செய்தால் மலேசியத் தமிழ் இணையம் பெரிதாக வளர்ச்சியடையும். அதற்குரிய மனித மூலதனம், ஆற்றல், வாய்ப்பு, வசதி ஆகிய அனைத்தும் மலேசியத்தில் நிரம்ப இருக்கின்றது.

மேற்கோள் பட்டி:-

1. தமிழ் இணையம் 2001 மாநாட்டுக் கட்டுரைகள், கோலாலம்பூர்
2. முரசு.நெடுமாறன், மலேசியத் தமிழரும் தமிழும், உலகத் தமிழாராய்ச்சி நிறுவனம்(2007), சென்னை, ப.250 – 264
3. தேவநேயப் பாவாணர் நூற்றாண்டு விழா, மலேசிய மலர்(2002), ப.271 – 273
4. <http://www.anjal.net/>
5. <http://www.treasurehouseofagathiyar.net/>
6. <http://tamilnetmalaysia.com/>
7. <http://www.malaysiaindru.com/>
8. <http://thirutamil.blogspot.com/>
9. <http://thirumandrill.blogspot.com/>
10. http://i140.photobucket.com/albums/r11/sathis_divine/valaipoongaa-2.png
11. http://jpnperak.edu.my/v2/modules/mastop_go2/go2.php?tac=17
12. <http://www.1malaysia.com.my/index.php?lang=ta>

12

**மின்னரசும் தமிழ் தகவல்
தொழில்நுட்பமும்**



மின்னாளுமை இணையச் செயலகம்

Albert Fernando,

Email: albertgi@gmail.com

எங்கும் தமிழ்;எதிலும் தமிழ் என்பதற்கொப்ப இணைய வெளியில் இப்போது தமிழ் அரசோச்சும் அற்புதம் அங்கிங்கெனாதபடி எங்கும் வியாபித்துள்ளது. இது இணைய வரலாற்றில் ஒரு பொன் பொழுது!

தமிழக அரசு அலுவலகங்களை காகிதங்கள் இல்லாத கணினியால் மட்டும் நிர்வகிக்கும் பணியினை இப்போதிருந்தே துவக்க வேண்டும். மின் ஆளுமை / மின்னாளுமை என்ற இ-கவர்னன்ஸ் இன்றைய காலத்தின் கட்டாயமான ஒன்றாகி விட்டது.

இன்று உலகம் தகவல் தொழில்நுட்பம் இன்றேல் எதுவுமில்லை என்ற நிலையை நெருங்கி வந்து கொண்டிருக்கிறது. தமிழ் நாட்டிலோ எண்ணிறந்த மக்கள் இன்னும் பழைய படிப்பறிவற்ற வாழ்க்கை நடை முறைகளையே பெரும்பாலும் கொண்டிருக்கின்றனர். அரசு அலுவலகங்களோ கோப்புகளை நிர்வகிக்கும் சடங்குகளையே கடைப் பிடிப்பதால் பொதுமக்கள் பலனைப் பெறுவதில் தாமதங்களே உயர் நிலைவகிக்கிறது.

தமிழகத்தில் 32 மாவட்டங்கள். முப்பதுக்கும் மேற்பட்ட அரசு மற்றும் அரசு சார்புத் துறைகள் உள்ளது. ஒவ்வொரு துறையும் தகவல் தொழில் நுட்பத்தால் இணைத்து ஒருங்கிணைந்த அரசுத்துறை "இணையச் செயலகம்" உருவாக்க வேண்டும். அப்படி உருவாக்கி தலைமைச் செயலகத்தோடு இந்த இணையச் செயலகம் தொடர்புபடுத்தப்படல் வேண்டும்.

ஒவ்வொரு மாவட்டத்திலும் எப்படி மாவட்ட ஆட்சித் தலைவர் அலுவலகம் இருக்கிறதோ, அப்படி மாவட்ட இணைய செயலகம் இயங்க வேண்டும். இந்த மாவட்ட இணையச் செயலகத்தோடு அந்த மாவட்டத்தில் உள்ள அத்துணை அலுவலகங்களும் இணைக்கப்பட வேண்டும் .

இந்த மாவட்ட இணைய அலுவலகங்கள் மாநிலத் தலைநகரில் தலைமைச் செயலகத்தோடு அல்லது தலை நகரின் ஒரு பகுதியில் "மாநில இணையச் செயலகம்" என்ற மின்னாளுமை இணையச்செயலகத்தோடு இணைக்கப்பட வேண்டும். ஏற்கனவே NIC எனப்படும் மத்திய அரசு அலுவலகம் மாவட்ட அளவில் இயங்குகிறதே என்று நினைக்கலாம். அதன் பணி வேறு;நான் சொல்லக்கூடிய இணையச் செயலக பணிவேறு. தமிழகத்தின் ஒரு குக்கிராமம்

முதல் தலை நகர் வரை இந்த இணையச் செயலகத்தில் இணைக்கப்பட வேண்டும். எப்படி?

குறிப்பிட்ட ஊராட்சிகள் அடங்கியது ஊராட்சி ஒன்றியமாகும். தமிழகத்தில் 385 ஊராட்சி ஒன்றியங்கள்! இந்த ஊராட்சி ஒன்றியங்களில் 12,618 ஊராட்சிகள் உள்ளன. ஊராட்சியில் பல உட்கடைக் கிராமங்கள் இருக்கும். தமிழக அரசு ஒவ்வொரு ஊராட்சிக்கும் ஒரு கணினியை வழங்கியுள்ளது. இந்தக் கணினி தான் இந்த இணையச் செயலகத்தின் அடிநாதம் என்றால் மிகையிலலை.

ஊராட்சிக்கு உட்பட்ட கிராமங்களின் தேவைகள், சாலை வசதிகள்,குடிநீர் வசதிகள், போக்குவரத்து, மின்சாரம் போன்ற பொதுத் தேவைகள் மற்றும் பொதுமக்களின் குடும்ப அட்டை,முதியோர் பென்சன், சமூக நலத்திட்டத்தின் பல்வேறு உதவிகள் பெற அந்த கிராமத்திலிருந்து சம்பந்தப்பட்ட அலுவலகத்துக்கு தங்கள் அன்றாட ஜீவனத்திற்கான கூலி வேலைகளைக் கூட விட்டுவிட்டு அலைகிறார்கள்.

அந்தந்த கிராம மக்கள் அந்தந்த ஊராட்சியில் தங்கள் மனுவை,புகாரைக் கொடுத்துவிட்டால் அது மாவட்ட இணையச் செயலகத்துக்கு அனுப்பப்படும். அங்கிருந்து சம்பந்தப்பட்ட துறை அலுவலக நடவடிக்கைக்கு அனுப்பப்படும். குறிப்பிட்ட மனுக்கள் குறிப்பிட்ட காலக்கெடுவிற்குள் முடிக்கப்பட்டு விட்டதா? என்பதையும் இந்த அலுவலகம் கண்காணிக்கும். பொதுமக்கள் புகாரில் சில மாநிலத் தலைநகரில் உள்ள

அலுவலகத்தால் நடவடிக்கை எடுக்கப்படவேண்டியது என்றால் அதை இந்த மாவட்ட இணைய அலுவலகம் கவனிக்கும். சரி.

ஊராட்சி,ஊராட்சி ஒன்றியங்கள் போல பேரூராட்சி,தேர்வு நிலை பேரூராட்சி 561, நகராட்சி,நகரியம், 8 மாநகராட்சி என்ற அலுவலக ஆளுகைகளும் இந்த இணையச் செயலக வட்டத்திற்குள்தான் வரும்!

மாவட்ட இணையச் செயலகம் அன்றாடம் மாவட்டத்தின் பல்வேறு துறைகளில் இருந்து தலைமைச் செயலகத்துக்கு,பிற துறைகளுக்கும் அனுப்பிப் பெறப்படவேண்டிய ஒப்புதல்கள், அங்கீகாரங்கள், தகவல் பரிமாற்றங்கள் அனைத்தையும் மண்ணஞ்சல் அல்லது அலுவலக பணியாளர் மூலம்

அனுப்பி பெறப்படும் செயல்களை மின்னஞ்சல்கள் மூலமாகவே இந்த இணையச் செயலகம் செய்யும்.

அதுமட்டுமல்ல சட்டமன்றம் நடக்கும் காலங்களில் குறிப்பிட்ட புள்ளிவிபரம் வேண்டும் என்றால் தலைமைச் செயலகத்தில் இருந்து ஒவ்வொரு மாவட்டத்தையும் தொடர்புகொண்டு புள்ளிவிபரங்கள் பெறும் நிலை மாறி தலைமைச் செயலகம் அமைந்துள்ள இடத்தில் அமையப்பெறும் இந்த மின்னாளுகை இணையச் செயலகத்தில் நொடியில் இந்தப்புள்ளி விபரங்களைப் பெற முடியும்.

முக்கியமாக அலுவலகங்களில் பராமரிக்கப்படும் கோப்புகள் இல்லாமல் போகும்;காகிதங்கள் கட்டுக் கட்டாக அடுக்கி காக்கப்படும் நிலையம் ஒழிந்துபோகும்.

இதனால் அரசுக்கு இடமும் மிச்சம்;பலகோடி ரூபாய் செலவினமும் மிச்சம். எப்படி என்பதற்கு ஒரு உதாரணம் மட்டும் சொல்கிறேன்.

உதாரணமாக ஒரு கிராமத்திலிருந்து பட்டா மாறுதல் கோரி மனு அனுப்புகிறார். ஆட்சியர் அலுவலகத்தில் பதிவு எழுத்தர் அந்த மனுவைப் பெற்றதும் அந்த மனுவைப் பெற்ற தேதி, மனு குறித்த சுருக்கமான விபரம் போன்றவற்றை அஞ்சல் பதிவேட்டில் பதிவு செய்துகொள்கிறார். பின், உரிய அலுவலருக்கு அந்த மனுவை எடுத்துக் கொண்டுபோய் அளிக்கிறார். அந்த அலுவலர் அவருக்குரிய பதிவேட்டில் பதிவு செய்து அதற்கென ஒரு கோப்பையும் அந்த மனுவுக்கு ஒரு எண்ணை உருவாக்கியும் வருவாய் ஆய்வாளருக்கு ஒரு கடிதம் எழுதி அந்தமனு குறித்து விசாரித்து அறிக்கை தரும்படி எழுதி சுருக்கொப்பமிட்டு, தன் அலுவல் மேலாளரின் கையொப்பத்துக்கு அனுப்புவார். மேலாளர் கையொப்பமிட்டு மீண்டும் அலுவலருக்கு வரும். அவர் அஞ்சல் அனுப்பும் பதிவேட்டில் மனு யாருக்கு அனுப்பப்படுகிறது என்ற சுருக்க விபரத்தை பதிந்து கொண்டு அஞ்சல் அனுப்புகை எழுத்தருக்கு அனுப்புவார்.

அஞ்சல் அனுப்புகை எழுத்தர், அந்த மனு யாருக்கு முகவரி இடப்பட்டுள்ளது என்று பார்த்து அதை அவருடைய அனுப்புகை பதிவேட்டில் பதிந்து உரிய அரசு அஞ்சல் வில்லை ஒட்டி அனுப்புவார். அந்த அஞ்சல் உரிய வருவாய் ஆய்வாளரைச் சென்றடைந்ததும் அவர் தமது பதிவேட்டில் கடித விபரச் சுருக்கத்தை பதிவு செய்து மனுதாரர் சம்பந்தப்பட்ட கிராம அலுவலருக்கு அந்த மனுவை விசாரணை செய்து அறிக்கை சமர்ப்பிக்குமாறு அனுப்புவார். அந்த கிராம நிர்வாக அலுவலர் அந்த மனுவை எடுத்துக்கொண்டு நேரிலோ அல்லது தலையாரி மூலமோ சம்பந்தப்பட்ட மனுதாரரை வரவழைத்து அவர் முதியோர் உதவித்தொகை பெற அருகதையுள்ளவரா? என்று சகல விதத்திலும் விசாரித்து தகுதியுடையவர்/தகுதியற்றவர் என்ற அவரது அறிக்கைப் பதிவை மீண்டும் வருவாய் ஆய்வாளருக்கு அனுப்புவார்.

எங்கிருந்து வந்ததோ அதே வழியில் மீண்டும் அந்த மனு பயணப்பட்டு முதலில் மனுதாரர் அனுப்பிய பிரிவுக்கே சென்று சேர்ந்து அங்கிருந்து தக்க உத்தரவு பிறப்பிக்கப்பட்டு மனுதாரரைச் சென்றடையும். இதில் எத்தனை அரசு அலுவலர்கள் அந்த மனுவிற்காக தங்கள் நேரத்தை செலவழிக்கிறார்களோ அது ம

னித உழைப்பாக கருதப்படுகிறது. அந்த மனித உழைப்பிற்காகும் மணித்துளிகள் பல்வேறு ஊதிய வேறுபாடுடைய அலுவலர்களின் வியர்வைத் துளிகள் கணக்கிடப்படுகிறது. சராசரியாக ஒரு மனு குறைந்தபட்சம் மூன்று வேலை நாட்கள் எடுத்துக்கொள்கிறது என்றால் 24 மணி நேரங்கள் ஆகும் என்று கருதப்படுகிறது. 24 மணி நேரத்துக்கு பல்வேறு அரசு அலுவலர்கள் அரசிடம் பெறுகிற ஊதியம் என்று கணக்கிடப்பட்டால் மணிக்கு சராசரியாக 300 ரூபாய் என்றால் தோராயமாக 7,200 ரூபாய் ஆகிறது.

ஒரு மாவட்டத்திலிருந்து நாளொன்றுக்கு சுமார் 400 மனுக்கள் பொதுமக்களிடமிருந்து பெறப்படுகிறது. மாதமொன்றுக்கு தோராயமாக பத்தாயிரம் என்றால் கூட வருடத்துக்கு ஒரு இலட்சத்து இருபதாயிரம் மனுக்கள். ஒரு மாவட்டத்திற்கு வருடத்துக்கு 86,40,000 ரூபாய்! தமிழகத்தில் 32 மாவட்டங்கள். 32 மாவட்டங்களில் மனுக்களை பெற்று நடவடிக்கை எடுக்க மட்டும் ஆகும் மனித உழைப்புக்கு அரசு (செலவளிக்கும்) அளிக்கும் ஊதியம் சராசரியாக ரூ.27,64,80,000 ஆகும்! ஏ! அப்பா!?

- மனு ஒன்றுக்கு 3 மனித உழைப்பு நாள்!
- மாதம் பத்தாயிரம் மனுக்கள் என்றால் 30,000 மனித உழைப்பு நாட்கள்!
- வருடத்திற்கு 3,60,000 மனித உழைப்பு நாட்கள்!
- 32 மாவட்டங்களுக்கு 21,52,0000 மனித உழைப்பு நாட்கள்!
- ஒரு மனுவிற்காக அலுவலகத்தில் பல்வேறு பிரிவுகளில் உபயோகப்படுத்தப்படும் காகிதப் பயன் பாட்டின் நிறை/எடை முப்பது முதல் 75 கிராம் வரை
- மாவட்டத்தில் மாதம் பத்தாயிரம் மனுக்களுக்காக உபயோகிக்கப்படும் காகிதத்தின் எடை தோராயமாக 7,50,000 கிராம்கள்.
- 32 மாவட்டங்களில் மனுக்களுக்காக உபயோகிக்கப்படும் காகிதத்தின் எடை தோராயமாக 22,65,0000 கிராம்கள்.
- ஒரு மனுவிற்காக ஒவ்வொரு அலுவலகத்திலும் செலவழிக்கப்படும் மொத்த அஞ்சல் வில்லைகளின் மதிப்பு ரூ.22/=

மாவட்டத்தில் மாதம் பத்தாயிரம் மனுக்களுக்காக செலவழிக்கப்படும் மொத்த அஞ்சல் வில்லைகளின் மதிப்பு ரூ.2,20,000/=

32 மாவட்டங்களில் மனுக்களுக்காக செலவழிக்கப்படும் மொத்த அஞ்சல்

வில்லைகளின் மதிப்பு ரூ.70,40,000

ஆனால், கணினியைப் பயன்படுத்தி இந்த வேலையைச் செய்யும் பொழுது பலமடங்கு குறைக்கப்படுகிறது. கோப்புகளில் பதிவது, அலுவலர், மேலாளர், உயர் நிலை அலுவலர் என்ற பல அடுக்கு முறைப் பணிகள் குறைக்கப்பட்டு சுருக்கப்பட்டு மனித உழைப்பு ஒரு மனுவிற்கு ஒருநாளாக்கப்படுகிறது. மூன்றில் ஒரு பங்கு மட்டுமே மனித உழைப்பு பயன்படுத்தப்படுகிறது! காகிதம் இல்லா கணினியில் மனுவின் மூலம் முதல் அதன் இறுதிச் செயற்பாடுகள் அனைத்தும் பதிந்து சேமிக்கப்படுகிறது; இதனால் மூன்றில் இரண்டு பங்கு அரசுக்கு செலவினம் குறைகிறது. அப்படிச் செல்வினம் குறைகிறது என்றால் அது இந்த அரசுக்கு சேமிப்பு! செலவளிக்காத சேமிப்பு அரசுக்கு அது வருவாய் அல்லாமல் வேறு என்ன?

நான் சொன்ன புள்ளிவிபரங்கள் மாநிலம் முழுவதும் உள்ள மாவட்ட ஆட்சித் தலைவர் அலுவலகங்கள் மட்டும்தான்! இதுபோல மாநிலத்தில் உள்ள 1,500க்கும் மேற்பட்ட அலுவலகங்களை இணைத்து ஒரு குடையின் கீழ் அனைத்து தகவல்களையும் கொண்டுவரும் செயற்பாட்டை இந்த மின்னாளுகை இணையச் செயலகம் மேற்கொள்ளும். உதாரணமாக கன்னியாகுமரியில் உள்ள ஒரு குக்கிராமப் பஞ்சாயத்திலிருந்து கடலோரத்தில் மண்ணரிப்பு ஏற்படுகிறது. இதனை மண் மூட்டைகள் அடுக்கி எங்கள் கிராமத்துக்கு பாதுகாப்பு ஏற்படுத்தவேண்டும் என்ற ஒரு மனு அந்தப் பஞ்சாயத்து தலைவரால்

புதன் கிழமை மாவட்ட ஆட்சியருக்கு அனுப்பப்படுகிறது. புதன்கிழமை மாவட்ட இணையச் செயலக அலுவலர் இந்த மனுவை ஆட்சியரின் கவனத்துக்கு கொண்டு செல்கிறார். ஆட்சியர் பிறப்பிக்கும் பரிந்துரையை சென்னையிலுள்ள மின்வளத்துறை அமைச்சகத்துக்கு அனுப்ப அன்றே ஆட்சியரின் பரிந்துரையை ஏற்று உரிய உத்திரவை மின்னஞ்சல் மூலமாக அனுப்ப அன்று மாலையே ஆட்சியர் உத்திரவைப்பெற்று தமது துறை அலுவலருக்கு குறிப்பிட்ட நிதியிலிருந்து மண் மூடைகள் வாங்கி உடனடியாக அந்தக் குக்கிராமத்தின் கடற்கரையில் மண்மூட்டை அணை கட்ட பிறப்பிக்க, மறுநாள் அந்தப்பணி அங்கே நிறைவடைகிறது. இப்படிச் செய்யும்போது தமிழகம் இணைய வானின்

மின்னாளுகைக்குள் வந்துவிடும்! இந்தியாவின் பிற மாநிலங்கள் வியந்து விழிமடல் திறக்கும் சூழலை உருவாக்கும்.

எதிர்கால சவால்களை சமாளிக்கும் வகையில், சமூகத்தை பன்முகம் கொண்டதாக மாற்றி அமைக்க வேண்டிய பொறுப்பும் கடப்பாட்டோடும் இதை திட்டமிட்டு உருவாக்க வேண்டும். அப்படி உருவாக்கும் போது ஒவ்வொரு துறையின் தகவல் பரிமாற்றம், திட்ட வரைவு சமர்ப்பித்தல், அனுமதி, செயலாக்கம் அனைத்தும் தமிழில் மின்னல் வேகத்தில் நடைபெறும். தாமதங்கள் தவிர்க்கப்படும்.

தமிழக அரசு ஒவ்வொரு துறையிலும் முத்திரையைப் பதிக்க வேண்டும். தகவல் தொழில் நுட்பத்தை முழுமையாகப் பயன்படுத்தினால் ஒவ்வொரு துறையும் சிறப்பாக இயங்க முடியும்.

- ஆல்பர்ட் ஃபெர்னாண்டோ, விசுகான்சின், அமெரிக்கா.

சான்றாதாரங்கள்:-

- http://thoduvaaanam.blogspot.com/2010_01_01_archive.html
- <http://namakkalcollector.net/>
- <http://www.tn.gov.in/districts.html>
- <http://www.tn.gov.in/departments.html>
- <http://panchayat.nic.in/index.do?siteid=101&sitename=Government%20of%20India%20
%20Ministry%20of%20Panchayati%20Raj>
- <http://panchayat.nic.in/viewMore.do?ppid=200&ptltid=375&itemid=1>

Prototype for E-government issued E-card using RFID technology based on Tamil font database

Vijayalakshmi.S.R

Assistant Professor, School of IT and Science,
Dr.GRD College of Science, Coimbatore-14, India.
srvijisiva@gmail.com

Abstract

RFID is not a new technology and has passed through many decades of use in military, airline, library, security, healthcare, sports, animal farms and other areas. Industries use RFID for various applications such as personal/vehicle access control, departmental store security, equipment tracking, baggage, fast food establishments, logistics, etc. The enhancement in RFID technology has brought advantages that are related to resource optimization, increased efficiency within business processes, and enhanced customer care, overall improvements in business operations and healthcare. Passports and other identification documents may be enhanced using the advancements in RFID technology. Various national and international bodies are pursuing machine-readable approaches with biometric information. This paper examines implementation regarding these electronic approaches and developments toward electronic data storage and transmission. Radio-frequency identification (RFID) devices for electronic passports and other existing identity documents are discussed. Our proposed research aim is to produce a model for e-government. All identity information about the citizen is included and database is maintained in Tamil language. Every citizen is having voter ID, PAN card, Credit card, ration card, driving License information, insurance information, Passport and many more. All cards can be replaced with single E-card. In the proposed model, database (all information about the citizen) is stored in tamil font using php mySQL database. This E-card provided by the government could be used by the citizen. The E-government maintains database in Tamil of all citizens. It helps in identifying the person during voting time. It can be used as security. However, the focus of this paper is to explore the main RFID components, i.e. the tag, antenna and reader.

Keywords: RFID technology, Electronic Passport, E-card, Biometrics.

1. Introduction

RFID stands for Radio Frequency Identification and is a term that describes a system of identification. RFID is based on storing and remotely retrieving information or data as it consists of RFID tag, RFID reader and back-end Database. RFID tags store unique identification information of objects and communicate the tags so as to allow remote retrieval of their ID. RFID technology depends on the communication between the RFID tags and RFID readers. The range of the reader is dependent upon its operational frequency. Usually the readers have their own software running on their ROM and also, communicate with other software to manipulate these unique identified tags. Basically, the application which manipulates tag deduction information for the end user, communicates with the RFID reader to get the tag information through antennas. Many researchers have addressed issues that are related to RFID reliability and capability. RFID is continuing to become popular because it increases efficiency and provides better service to stakeholders. RFID technology has been realized as

a performance differentiator for a variety of commercial applications, but its capability is yet to be fully utilized.

2. RFID Evolution

RFID technology has passed through many phases over the last few decades. The technology has been used in tracking delivery of goods, in courier services and in baggage handling. Other applications includes automatic toll payments, departmental access control in large buildings, personal and vehicle control in a particular area, security of items which shouldn't leave the area, equipment tracking in engineering firms, hospital filing systems, etc.

3. How RFID System Works

Most RFID systems consist of tags that are attached to the objects to be identified. Each tag has its own "read-only" or "rewrite" internal memory depending on the type and application. Typical configuration of this memory is to store product information, such as an object's unique ID personal details, etc. The RFID reader generates magnetic fields that enable the RFID system to locate objects (via the tags) that are within its range. The high-frequency electromagnetic energy and query signal generated by the reader triggers the tags to reply to the query; the query frequency could be up to 50 times per second. As a result communication between the main components of the system i.e. tags and reader is established.

If the reader is on and the tag arrives in the reader fields, then it automatically wakes-up and decodes the signal and replies to the reader by modulating the reader's field. All the tags in the reader range may reply at the same time, in this case the reader must detect signal collision (indication of multiple tags). Signal collision is resolved by applying anti-collision algorithm which enables the reader to sort tags and select/handle each tag based on the frequency range (between 50 tags to 200 tags) and the protocol used. In this connection the reader can perform certain operations on the tags such as reading the tag's identifier number and writing data into a tag. The reader performs these operations one by one on each tag.

4. Components of an RFID System

This RFID system allows to deduct the objects (tag) and perform various operations on it. The integration of RFID components enables the implementation of an RFID solution. The RFID system consists of following five components (as shown in Figure 1):

- Tag (attached with an object, unique identification).
- Antenna (tag detector, creates magnetic field).
- Reader (receiver of tag information, manipulator).
- Communication infrastructure (enable reader/RFID to work through IT infrastructure).
- Application software (user database/application/ interface).

5. Tags

Tags contain microchips that store the unique identification (ID) of each object. The ID is a serial number stored in the RFID memory. The chip is made up of integrated circuit and embedded in a silicon chip. RFID memory chip can be permanent or changeable depending on the read/write characteristics. Read-only and rewrite circuits are different as read-only tag contains fixed data and can not be changed without re-program electronically. On the other hand, re-write tags can be

programmed through the reader at any time without any limit. For example, in the case of the credit cards, small plastic pieces are stuck on various objects, and the labels. Labels are also embedded in a variety of objects such as documents, cloths, manufacturing materials etc.

There are two types of tags (active and passive) are being used by industry and most of the RFID system. The essential characteristics of RFID tags are their function to the RFID system. This is based on their range, frequency, memory, security, type of data and other characteristics. These characteristics are core for RFID performance and differ in usefulness/support to the RFID system operations.

5.1 Tag Frequencies

The range of the RFID tags depends on their frequency. This frequency determines the resistance to interference and other performance attributes. The use/selection of RFID tag depends on the application; different frequencies are used on different RFID tags. The following are the commonly used frequencies:

- **Microwave** works on 2.45 GHz, it has good reader rate even faster than UHF tags. Although at this frequency the reading rate results are not the same on wet surfaces and near metals, the frequency produce better results in applications such as vehicle tracking (in and out with barriers), with approximately 1 meter of tags read range.
- **Ultra High Frequency** works within a range of 860-930 MHz, it can identify large numbers of tags at one time with quick multiple read rate at a given time. So, it has a considerable good reading speed. It has the same limitation as Microwave when is applied on wet surface and near metal. However, it is faster than high frequency data transfer with a reading range of 3 meters.
- **High Frequency** works on 13.56MHz and has less than one meter reading range but is inexpensive and useful for access control, items identifications on sales points etc as it can implanted inside thin things such as paper.
- **Low Frequency** works on 125 kHz, it has approximately half a meter reading range and mostly used for short reading range applications

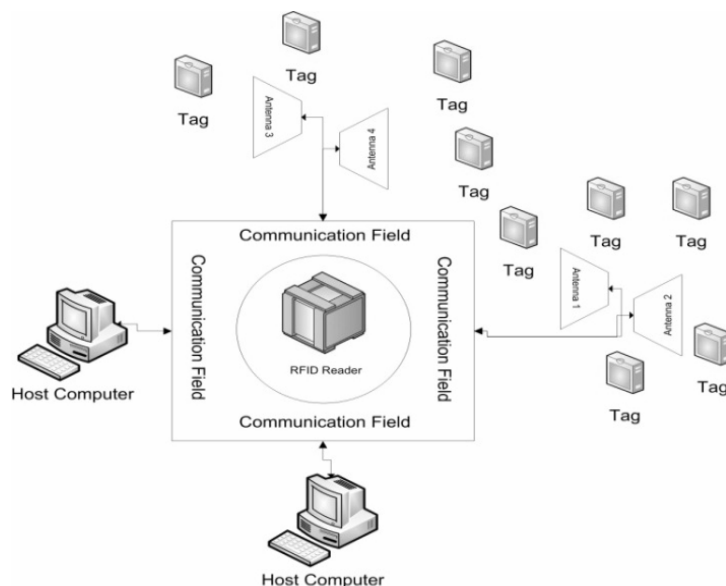


Fig 1 Components of an RFID System

6. Antennas

RFID antennas collect data and are used as a medium for tag reading. It consists of the following: (1) Patch antennas, (2) Gate antennas, (3) Linear polarized, (4) Circular polarized, (5) Di-pole or multi-pole antennas, (6) Stick antennas, (7) Beam-forming or phased-array element antennas, (8) Adaptive antennas, and (9) Omni directional antennas.

7. RFID Reader

RFID reader works as a central place for the RFID system. It reads tags data through the RFID antennas at a certain frequency. Basically, the reader is an electronic apparatus which produce and accept a radio signals. The antennas contains an attached reader, the reader translates the tags radio signals through antenna, depending on the tags capacity. The readers consist of a build-in anti-collision schemes and a single reader can operate on multiple frequencies. As a result, these readers are expected to collect or write data onto tag (in case) and pass to computer systems. For this purpose readers can be connected using RS-232, RS-485, USB cable as a wired options (called serial readers) and connect to the computer system. Also can use WiFi as wireless options which also known as network readers. Readers are electronic devices which can be used as standalone or be integrated with other devices and the following components/hardware into it. (1) Power for running reader, (2) Communication interface, (3) Microprocessor, (4) Channels, (5) Controller, (6) Receiver, (7) Transmitter, (8) Memory.

8. Storing Tamil font using PHP:

By default mysql supports many european languages, Since unicode character (UTF-8) support implemented in mysql it allows us to store many of the indian (Asian) languages. Mysql supports Gujrathi, Hindi, Telugu and TAMIL among too many languages in the subcontinent. Let's consider TAMIL language and workout:-

To store & search tamil character sets in MySQL table, first of all we need to create a table with character set UTF-8.

```
CREATE TABLE multi_language
(
id INTEGER NOT NULL AUTO_INCREMENT,
language VARCHAR(30),
characters TEXT,
PRIMARY KEY(id)
) ENGINE=INNODB CHARACTER SET = utf8;

INSERT INTO multi_language VALUES (NULL, 'English', 'welcome');
INSERT INTO multi_language VALUES (NULL, 'Arabic', 'نم لك ي ط ح ز و ه د ج ب أ');
INSERT INTO multi_language VALUES (NULL, 'Arabic', 'ن م ل ك ي ط ح ز و ه د ج ب أ');
INSERT INTO multi_language VALUES (NULL, 'Tamil', 'Tamil character letters');
```

If the result shows like the???? then properly in the system (windows XP) need to install the extra language support tool by enabling the following options,

Control Panel -> Regional and Language Option -> Languages -> Install files for complex script and right-to-left languages (including Thai)

While fetching the row using PHP, for displaying the Multilanguage content properly it is needed to include the Meta tag like,

```
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=utf-8">
```

if needed the add `mysql_query('SET character_set_results=utf8')` in the php code before fetching the record. Using this software the database could be created in tamil for the citizen information. i.e.E-card information. E-card information includes personal information, driving license, medical policy, PAN number, Passport number, voter ID, Credit card, Ration card and many more information in Tamil language.

11. Conclusions

This study has identified and explained the nature of RFID technology evolution with respect to RFID applications. RFID technology will open new doors to make organizations, companies more secure, reliable, and accurate. The first part of this paper has explained and described the RFID technology and its components, and the second part has discussed the main considerations of Tamil font in the PHP my SQL. The paper considers RFID technology as a means to provide new capabilities and efficient methods for e-governance. The implementation of e-card has not been without challenges, and some continue to challenge the use of contactless technology and other identity documents. This paper analyzed the major current and potential uses of RFID in identity documents. This paper also gave a model for E-card using RFID technology and E-card information could be stored in the tamil font databse.

References

- 1) Yingjiu Li, Member, Vipin Swarup, and Sushil Jajodia, Fingerprinting Relational Databases:Schemes and SpecialtiesIEEE, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 2, NO. 1, JANUARY-MARCH 2005 pp 34 to 45.
- 2) S. Garfinkel, B. Rosenberg, "RFID Application, Security, and Privacy", USA, (2005), ISBN: 0-321-29096-8.
- 3) L. Srivastava, RFID: Technology, Applications and Policy Implications, Presentation, International Telecommunication Union, Kenya, (2005).
- 4) A. Narayanan, S. Singh & M. Somasekharan, "Implementing RFID in Library: Methodologies, Advantages and Disadvantages", (2005).
- 5) S. Shepard, (2005), "RFID Radio Frequency Identification", (2005), USA, ISBN:0-07-144299-5.

Recommended Approach for e-Governance in Tamil Nadu using MPMLA.IN

Syed Hussain

(admin@auroinfotech.com)

Core Objective

The topic of this presentation is based on my experiences with MPMLA.IN website <http://mpmla.in/>. The objective is to present this working model as a sample way to implement eGovernance using a mixture of technologies related to Tamil language to help the elected representatives to interact with general public and also to use it to effectively improve the quality of life of common man.

Current Implementation and Limitations

Currently, MPMLA.IN covers all the MPs and MLAs in every single parliamentary and assembly constituency in India. The website is an unbiased effort to facilitate communication between people and their respective MPs/MLAs in India. The people can put across their grievance/appreciation to their MP/MLA and the MP/MLA can choose to respond and address the issue. There is no bias or political affiliation.

Key Features:

Some of the key features available for the visitors to MPMLA.IN today are:

- Easy identification of MPs and MLAs from the website by State, Parliamentary Constituency and Assembly Constituency.
- Search by MP, MLA, Party, Constituency or Politician.
- Easy entry of problems using English language as well as typing in English automatically converted to Tamil.
- Every feedback and problem reported is validated by moderators before publishing on the site.
- Top Five Issues List: A Top Five Issues list is available across each constituency thus helping the elected representatives to focus on the core problems reported in their constituency.

Major Limitations

Some of the main limitations we identified are:

- People who are not comfortable in English are forced to express their thoughts in English.
- People who know English and Tamil try to type in Tamil language using English alphabets.
- People who only know Tamil cannot express their concerns on the website
- People who do not have a computer cannot express their concerns

The idea behind this paper is to recommend an enhanced version of MPMLA.IN as a viable model for e-Governance using various technology advances in Tamil. This model also addresses the major limitations mentioned above, along with many additional features. The following page has a screen

shot of how MPMLA.IN looks today. This is followed by sections explaining the new technology advancements recommended in the near future and also the long term plan for the same.

Screen shot of MPMLA.IN website as of 14-Apr-2010

MPMLA MP/MLA Constituency Party Politician **SEARCH**

HOME | FEED BACK | TELL A FRIEND | CREDITS | TERMS OF SERVICE | PRIVACY POLICY

» MAHARASHTRA » NATIONAL CAPITAL TERRITORY OF DELHI » TAMIL NADU » ANDHRA PRADESH » GUJARAT

Tamil Nadu : Mayiladuturai : Kuttalam : Anbazhagan Kuthalam K. : DEAR MLA, MY HEARTIEST CONRATULATIONS TO OUR ONRIYA PERUND THALAIVAR MR. KARUNANIDHI, M., C.M.,

KARUNANIDHI, M. , C.M ,
TAMIL NADU, India

Capital: Chennai
Chief Minister (CM): KARUNANIDHI, M.

Number of MLAs: 235
Number of MPs: 39

Parliamentary Constituencies

Chennai North	T K S Elangovan
Chennai Central	Dayanidhi Maran
Chennai South	Rajendran
Sriperumbudur	Baalu, T. R.
Arakkonam	Jagathrakshakan, S. Dr.
Vellore	M.Abdul Rehman
Cuddalore	Azhagiri
Chidambaram	Thirumavalavan
Dharmapuri	Tamaraiselvan
Krishnagiri	Sugavanam, E. G
Salem	Seririvalei
Nilgiris	Raja A
Coimbatore	P.R.Natarajan
Pollachi	Sukumar
Dindigul	Chiththan, N. S. V.
Madurai	M K Azhagiri
Karur	Thambi Durai.M
Tiruchirappalli	P Kumar
Perambalur	T Nepolian
Mayiladuturai	O.S Marian
Nagapattinam	Vijayan, A. K. S.
Thanjavur	Palanmanickam, S. S.
Sivaganga	Chidambaram, P.
Ramanathapuram	K Siva Kumar Alias J K Ratheesh
Tirunelveli	Ramasubbu
Tenkasi	Lingam
Theni	Aarun
Kallakurichi	Adi Sanker
Viluppuram	M.Anandan
Arani	Krishnasswamy, M.
Tiruppur	Sivasamy
Virudhunagar	Manick Thakoor
Kanniyakumari	J Hellan Davidson
Namakkal	Gandhiselvan.S
Thoothukkudi	S R Jayadurai
Tiruvannamalai	T Venugopal
Tiruvallur	Dr.P.Venugopal
Erode	Ganesamoorthy
Kancheepuram	P . Viswanathan

Ask the CM of Tamil Nadu

Have any questions to ask the CM? Please enter it below!

Do not display this comment on the site

E-Mail Id

Phone Number

Code

<< first < prev 1 2 3 4 5 6 7 8 9 10 next > last >>

Past Feedbacks/Questions from users

i am an engineering student studying in chennai from thanjavur you are my role model sir . you are great person and taking tamilnadu and people of tamilnadu to the

Latest News

Retrieving RSS feed(s)

Photo Gallery

Secretariat
Tamilnadu Secretariat

Planned Implementation for eGovernance using MPMLA.IN

The following section gives a full scope of how the existing model can be adopted in Tamil Nadu by leveraging technology advancements in Tamil.

Overview

MPMLA.IN can be adopted by Government of Tamil Nadu as the central system for general public to report and track their day to day non-emergency issues. This system will facilitate better tracking and monitoring of any problem across the state.

The following are some of the core features to be integrated in the immediate future in MPMLA.IN:

Elected Official Login Management

- **Main Login Module:** Every elected official will be able to access the problems reported by the visitors using various methods such as computer, mobile devices, etc.
- **Communication Module:** The elected official can use this method to communicate their thoughts and ideas over a period of time.
- **Delegation:** Elected officials can also delegate their access to their subordinates – this way they do not have to always be present to type all the responses. The responses can be typed up by the delegates and submitted. This also makes it easy for the elected representative to quickly give follow ups based on each issue.

Problem Management Process

- Feedback and Problem Submission
- Computer – Typing in Tamil or English – Any one can access the MPMLA.IN website and type their problem in Tamil or English and report the issue.
- SMS – in Tamil or English – Any person can SMS their issues to a number from their mobile device.
- Voice Mail – in Tamil or English – The system allows users to call a toll free number and record their messages. This message will be automatically converted into text using a voice to text software.
- Picture – Users of the system can also choose to upload any photo that represents a problem better. For eg. If a road is in a poor condition, a visitor can take a photo and upload it from their mobile device or from a computer. This image will be associated with the problem mentioned by the visitor.

Problem Command Centre

A Problem Command Centre has to be established in order to review, analyse and also route the problem to its right officials. The Problem Command Centre also takes care of the following items:

- **Feedback Parsing:** Every feedback is automatically parsed and allocated based on the following parameters:
 - ❖ MP or MLA under whom the problem was reported
 - ❖ Area under which the problem was reported
 - ❖ Nature of the issue (eg. Road, Water, etc)
- **Problem Number Allocation:** Every feedback identified as a problem is allocated a unique Problem Number and tracked on the website. This Number helps track every issue from initiation to closure.
- **Service Level Agreement (SLA) with Escalation Procedure**
 - **SLA Time commitment:** Every elected representative will commit to a specific time within which an issue will be replied to. This commitment will be officially displayed on the site and

communicated to the problem reporter. Whenever a reply is made by any official, it will include the following:

- ❖ The current status of the issue
- ❖ Description of what has been done so far and what is planned
- ❖ When will the next update be provided on this issue
- ❖ Escalation Procedure: A well defined escalation procedure will be implemented in place which says who will be contacted if a person does not respond on time. This will make sure that if the issue crosses its SLA time, it will be escalated to the next higher official up in the chain of command. This will continue until the problem is addressed and resolved.

- **Periodic Analysis and Reporting**

- ❖ **Dashboards:** The plan is to provide extensive dashboard for each constituency so that the entire details about each issue are available at any time to any one visiting the constituency page. This also helps each official to understand key issues, their categories, etc
- ❖ **Reports:** Various reports will be generated to show problems that are unattended for a long time, repeated patterns, etc. This can be generated for any given time period, thus helping us to understand the history of an issue and identify repeatable pattern of issues, so that we can find a permanent fix for the same.
- ❖ **Comparison Charts:** Comparison Charts help us to understand how one constituency is handled with respect to another and how each elected official has performed over a period of time. This kind of metrics provide competitive analysis, based on which we can even arrive at the Constituency Index which will help us determine how a constituency fares in comparison to a similar constituency anywhere else in India.

Benefits for Elected Representatives

- **Additional Communication Medium:** This platform provides the elected representative an additional communication medium to interact with the citizens.
- **Transparent View of Status:** This platform provides the responsible officials with a full view of all the issues, their status and also how they are performing when compared to their peers.

Benefits for Citizens

- **Suggestion Box at Home:** People do not have to leave their homes and can directly report the feedback or issues using their computer or their cell phones.
- **Transparency and Accountability:** This model gives complete transparency on each issue reported by the people. This also improves the accountability on the part of the officials as they are aware that their performance is tracked publicly by every one.
- **Social Awareness and Involvement:** This form of interaction will automatically improve social awareness and improve the level of involvement for citizens in contributing towards improving their areas.

Future Enhancements

All Elected Representatives

One way of expanding MPMLA.IN would be to include the elected representatives at all levels starting from local municipality elections. This way every single elected representative will be able to have a full scope of their local issues as well as how it rolls up under the larger platform. This also provides full transparency and accountability at all levels.

All Government Officials

A second way of expanding MPMLA.IN would be to include all the Government officials belonging to various departments such as Public Works, Transportation, Industries and Commerce, etc. This will make sure there is complete transparency across all the departments that handle the day to day affairs of the people.

Conclusion

This approach is recommended to provide the simplest approach for any one in Tamil Nadu to leverage this system in order to have better reach to their elected representatives. A proper execution of this plan will ensure that there is better progress in every single area where there is a lack of attention today – thus leveraging various technologies to solve day to day problems for the common public.

References

- Election Commission of India Website
- Google Maps
- Google Transliteration Tool
- Vlingo Speech to Text Software
- Elgg Social Networking Software

சிங்கையில் நான் அனுபவித்த தமிழ்த் தொழில்நுட்ப வளர்ச்சி



திருமதி மீனாட்சி சபாபதி

ஆசிரியர், நூலாசிரியர், முன்னாள் மூத்த ஒலிபரப்பாளர்,

தன்முனைப்புப் பேச்சாளர்.

சிங்கப்பூர்,

Tholkaappiyam.blogspot.com

கணினி வளர்க்கும் தமிழ் எனும் மையக் கருத்துக்கு உட்பட்ட அனுபவக் கட்டுரை

ஒரு பாணை சோற்றுக்கு ஒரு சோறு பதம் என்பர். அந்த வகையில் எனது அனுபவம் சிங்கையில் தமிழ் தொழில்நுட்ப வளர்ச்சியை ஓரளவு அளந்து காட்டும் என நம்புகிறேன். ஏனெனில் காலத்தோடு இணைந்து வாழவும் புதிய தொழில்நுட்பங்களைப் பயன்படுத்தவும் தேவை உள்ள ஊடகத் துறையில் 22 ஆண்டு காலம் பணி புரிந்த அனுபவம் உள்ளவள் நான். அதன்வழி தமிழ் மொழியையும் பண்பாட்டையும் பரப்பும் பெரும் பாக்கியமும் பெற்றுள்ளேன். இந்த அனுபவ வரலாறு இந்த மாநாட்டின் வழி உலகத் தமிழரிடம் பதிவு செய்யப் படுமானால் அது வருங்காலத் தமிழர்க்கு ஒரு தகவலாக இருக்கும்.

கடந்து வந்த பாதை...

அன்றாட வாழ்க்கையில் அனைவரும் பயன்படுத்தும் பொருளாக இன்று கணினி விளங்குகிறது. சின்னப் பிள்ளைகள் எல்லாம் விருப்பம்போல் ஒளிபடம் (video) பிடித்து அதனை விரும்பியபடி மாயக் கோலங்கள் செய்து தொகுத்து படைக்கும் வண்ணம் தொழில் நுட்பக் கல்வி முறை இப்போது சிங்கப்பூர் போன்ற வளர்ந்த நாடுகளில் கிடைக்கிறது. மாணவர்கள் முதல் பெரியவர்கள் வரை அனைவரும் பயன்படுத்தும் இந்த சிறந்த சாதனங்கள் தமிழ் மொழி வளர்ச்சியோடு இணைந்து செயலாற்றத்தொடங்கியிருப்பதை என்னைப் போன்ற தமிழ் மொழி பரப்பாளர்கள் கண்டும் பயன்படுத்தியும் வருகிறோம். இது தொடர்பான என்னுடைய அனுபவத்தை சுருக்கமாக பகிர்ந்து கொள்கிறேன்.

1980 - களில் நான் சிங்கப்பூர் தேசிய பல்கலைக் கழகத்தில் படித்தபோது கணினித் துறை புதிதாக அறிமுகமானது. கணினி திட்டமுறைகளும் பயனீடுகளும் (Computer Programming and Applications) என்ற பாடத்தை அப்போது எடுத்துப் படித்த நாங்கள் punch card எனும் அட்டையை பயன்படுத்தி கணினிக்கு செயல் திட்ட உத்தரவு கொடுப்போம். Keyboard- விசை பலகைக் கருவி கண்டுபிடிக்கப்படாத காலம் அது. விசைப் பலகை வந்தபின் அதில் தமிழ் உருக் கொடுப்பது பற்றி சிங்கப்பூரில் பெரும் பேச்சு நிலவியது. நிறைய கலந்துரையாடல்கள், விவாதங்கள் இருந்தன. பலரும், ஆங்கில விசைக் கருவிக் கேற்ப தமிழில் எழுத்துக்களைக் குறைக்க வேண்டும் என்று வாதிட்டனர். அது உண்மையில், செருப்புக்கேற்ற அளவு காலை வெட்டுவதாகும் என்று பின்னர் புரிய வந்தது.

(இன்றும் கூட சிலர் தமிழ் எழுத்து சீர்திருத்தம் பற்றிப் பேசி வருவதை நாம் அறிவோம். தமிழின் பெருமையை மக்கள் மனதில் பதிய வைப்பதும் அதன் வழி ஆர்வத்தை மூட்டி தமிழ் படிப்போர் எண்ணிக்கையை அதிகரிப்பதும் தான் உண்மையான தேவை என்பதை அந்த சிலர் உணர வேண்டும்.)

சிங்கப்பூர் வானொலிப் பணியில்..

வானொலியைப் பொறுத்தவரை அந்தக் காலத்தில் மிகப் பிரபலமாக இருந்தது கடிதம் வழி நேயர் விருப்பம். நேயர் விருப்ப அட்டைகள் கடைகளில் விற்கப்பட்டன. நேரடி தொலைபேசி அழைப்புகள் வழி நேயர்கள் வானொலிக்கு அழைத்துப் பேசி ,பாடல் விரும்பிக் கேட்பது , சின்னப் பிள்ளைகள் கதை சொல்வதும் பாடல் பாடிக் காட்டுவதும் திருக்குறள் ஒப்புவிப்பதும் என்று பலவகையில் நிகழ்ச்சிகள் நடத்தப்படுகின்றன. இதனால் பேச்சுத் தமிழ் வளர்கிறது என்றால் மிகையிலலை. இடையில், தொலைப்பிரதி கொஞ்சம் தலைகாட்டியது. குறுந்தகவல் அறிமுகமாகி, பின் மின்னஞ்சல் வந்தது. அது பெரும் புரட்சி செய்தது.

1990களில் எங்கள் வானொலி 24 மணி நேர சேவை வழங்கத் தொடங்கியபோது மின்னஞ்சல், உலகத்தையே எங்களோடு இணைத்தது என்றுதான் சொல்லவேண்டும்.. நான் அந்த வேளையில் இரவு நேரப் பணியில் ஈடுபட்டிருந்தேன். பின்னிரவு இரண்டு முதல் அதிகாலை ஆறு வரை தினமும் வானொலியில் நிகழ்ச்சி படைத்தேன். அதில் குறிப்பாக அமெரிக்காவில் வாழும் நேயர்கள் அதிகம் பேர் கலந்து கொண்டனர். அந்த நேரத்தில் அமெரிக்காவில் பகல் வேளை. அப்போது கட்டுப்பாடின்றி இணைய ஒலிபரப்பு அனைவருக்கும் கிடைத்த சமயம். இப்போதுள்ளதுபோல் அன்று அங்கு தனியார் வானொலிகள் இருக்கவில்லை. ஆகவே சிங்கை வானொலியை அவர்கள் பயன்படுத்திக் கொண்டனர்.

வானொலியில் நான் அப்போது படைத்த அறிவோமா நாம் (தமிழர் பண்பாடும் இலக்கியமும் தற்கால அறிவியல் கண்டுபிடிப்புகளோடு எவ்வாறு பொருந்துகின்றன என்பது குறித்த ஆராய்ச்சித் தொடர்.) தமிழ் அமுது (எளிய தமிழ்ச் சொற்களை எவ்வாறு நாம் உரையாடலில் பயன்படுத்தலாம் என்பதை விளக்கும் உரையாடல் நிகழ்ச்சி), நாளும் ஒரு குரல் (தினமும் ஒரு குறட்பாவுக்கு, நடப்பு விவகாரங்களை மேற்கோள் காட்டி எளிய விளக்கம் தரும் நிகழ்ச்சி.) முதலிய நிகழ்ச்சிகளுக்கு பெருத்த வரவேற்பு கிட்டியது. தமிழ் மொழி மீது மிகுந்த ஆர்வம் கொண்டிருந்த அவர்களுக்கு நான் படைத்த 'அன்றாட வாழ்வில் தமிழ் இலக்கியப் பயனீடு' குறித்த நிகழ்ச்சிகள் மகிழ்ச்சியை அளித்ததில் வியப்பில்லைதான். அமெரிக்க நேயர்கள் பலர் மின்மடலில் தங்கள் பங்களிப்பை செய்தனர். தமிழ் மொழி மீது பற்றும் திறனும் வாய்ந்த அமெரிக்கத் தமிழர்க்கு மற்ற நாட்டுத் தமிழருடன் தொடர்பு கொள்ளவும் தமிழ்த் தாகத்தை தணிக்கவும் மிக சிறந்த வாய்ப்பாக அப்போது எங்கள் சிங்கப்பூர் வானொலி 'ஒலி 96 .8 ' இருந்தது. அந்த சமயத்தில் 'நாம்' என்ற எனது நிகழ்ச்சிக்காக அவர்கள் பலர் தங்களிடமிருந்த தமிழ்த் தகவல்களைப் பகிர்ந்து கொண்டனர்.

அமெரிக்கத் தமிழர்களின் இந்த ஈடுபாட்டுக்குக் கொஞ்சமும் சளைக்காமல் மலேசியத் தமிழர்களும் பங்கேற்றனர். வீட்டில் பிள்ளைகள் ஆங்கிலம் பேசுவதைத் தவிர்த்து தமிழ் பேச வைப்பது எப்படி என்பது முதல் திருமந்திரச் செய்யுள்கள் வழி எப்படி அறிவியல் ஆராய்ச்சியை மேம்படுத்தலாம் என்பது வரை பல்வேறு நிகழ்ச்சிகளை நான் படைத்துள்ளேன். என்னுடைய ஆய்வுகளும் முயற்சிகளும் தமிழ் நாடு, மலேசியா, ஆஸ்திரேலியா, மியன்மார், இங்கிலாந்து, அமேரிக்கா முதலிய பல நாடுகளில் உள்ள மக்களுக்கு சென்று சேர்ந்தது. பல நாடுகளிருந்தும் அழைப்புகள் வந்தன. எனது நூல்கள் பல நாடுகளுக்கும் பரவின. இணையம், மின்னஞ்சல் போன்றவை காரணமாகத்தான் எனக்கு அமெரிக்க தமிழ்ச் சங்கத்துடன் தொடர்பு கிடைத்தது. 2004 Fetna மாநாட்டில் நான் பங்கேற்றது அதனால்தான்.

இப்போது இணைய காப்புரிமை சட்டங்கள் காரணமாக சிங்கப்பூர் வானொலி ஒலிபரப்பு பல நாடுகளுக்குக் கிடைக்காமல் போனது வருத்தமே. வருத்தமில்லாத விஷயம் என்னவென்றால் You tube, blogging போன்றவற்றின் வளர்ச்சி காரணமாக இப்போது பல்வேறு இணையப் பக்கங்கள் தமிழில் கலக்குகின்றன. தமிழ்த் திரைப்படங்கள் , பாடல்கள் முதற்கொண்டு, தனிப்பட்ட நிகழ்ச்சிகளில் சிறுவர், இளையர், பெரியவர் பங்கேற்ற தமிழ்ப் படைப்புகள் வரை இணையத்தில் அனைவரின் பார்வைக்கும் வைக்கப்பட்டுள்ளன. மாபெரும் உலகத் தமிழ் சமுதாயத்திடம் சிங்கப்பூர் தமிழ் நடவடிக்கைகள் பற்றி எல்லா விவரங்களும் சென்று சேர முடிகிறது.

வானொலிக்கு அப்பால் ...

தற்போது நான் வானொலியில் பணியாற்றவில்லை என்றாலும் தொடர்ந்து எனது நேயர்களுடன் தமிழைப் பகிர்ந்து கொள்ள தகவல் தொழில்நுட்பம் வகை செய்கிறது. வானொலியில் இருந்தபோதே முடிந்த அளவு தமிழ் இணையப்பக்கங்களை படிக்கத் தூண்டியும் இருக்கிறேன். என் சொந்த வலைப் பூவையும் சேர்த்துதான். 'ஒலி 96 .8 ' இணையப் பக்கத்தோடு சேர்த்து எங்கள் படைப்பாளர்களின் வலைப்பக்கங்கள் இணைக்கப்பட்டிருக்கும். மற்ற வானொலி படைப்பாளர்கள் பெரும்பாலும் பொழுது போக்கு செய்திகள் வெளியிட்ட வேளையில் நான் தமிழ் தொடர்பான சீரிய செய்திகளையே வெளியிட்டேன்.

எடுத்துக்காட்டுக்கு தமிழ்ப் புத்தாண்டு தொடர்பான செய்திகள் குறித்து குழப்பங்கள் நிலவியபோது அது தொடர்பாக எனக்குக் கிடைத்த தகவல்களைத் திரட்டி என் வலைப்பதிவுப் பக்கமான tholkaappiyam.blogspot.com எனுமிடத்தில் வெளியிட்டேன்.

எ.கா. 2009-ஆம் ஆண்டுப் பதிவின் ஒரு பகுதி : அன்புள்ள நேயர்களே, தை முதல் நாள்! தமிழ்ப் புத்தாண்டு, பொங்கல், தமிழர் திருநாள் : இப்படி மூன்று பொருளில் சிறப்பு பெற்றுள்ள ஒரே நாள். இது குறித்து உங்களில் சிலர் கேட்ட கேள்விகளுக்கு உரிய பதில்களை இங்கு தந்திருக்கிறேன்.

இது குறித்த உங்களின் கருத்துகளையும் பகிர்ந்து கொள்ளுங்கள். உங்களிடம் கேள்விகள் வேறு ஏதேனும் இருந்தால் அதனையும் இந்த வலைப்பதிவில் கேளுங்கள். முடிந்த அளவு பதிலளிக்கிறேன். நன்றி. பொங்கல், புத்தாண்டு வாழ்த்துகளுடன், மீனாட்சி சபாபதி

- 1) தமிழ்ப் புத்தாண்டு நாள் எது? சித்திரை முதல் நாளா தை முதல் நாளா? இரண்டு நாட்களுமே புத்தாண்டாக பல காலம் கொண்டாடப்பட்டு வருகின்றன. தை முதல் தேதி பொங்கல் எனும் அறுவடைத் திருநாளாக விளங்குகிறது. தமிழர் வாழ்வில் இது முக்கியமான புத்தாண்டாக விளங்குவதால் இந்த நாளை புதுத் துணி அணிந்து புதுப் பானையில் சோறு பொங்கி விழாவாகக் கொண்டாடுவர். தை பிறந்தால் வழி பிறக்கும் என்பதையும் புது வருடப் பழமொழியாகக் கொள்வர். மேலும், புத்தாண்டுக்கு ஏதுவாக பழையனவற்றைக் கழித்துக் கட்டும் செயல் பொங்கலுக்கு முதல் நாள் மேற்கொள்ளப்படும். வேண்டாதவற்றைப் போக்கி விடும் நாள் போகி எனப்படுகிறது. இது, புது வருடச் செயலாக பல்லாண்டு காலம் தமிழர் வாழ்வில் இருந்து வரும் பழக்கம்.
- 2) இதுதான் புத்தாண்டு என்றால், சித்திரை முதல் தேதி ஏன் தமிழ்ப் புத்தாண்டாக அனுசரிக்கப் படுகிறது? சித்திரை மாதம் தொடங்கும் நாள்காட்டி முறை தமிழகத்தை பல்லவர் ஆண்ட காலத்தில் பிரபலமானதாக தகவல்கள் கூறுகின்றன. இந்த நாள்காட்டி முறையில் ஆண்டுகள் சுழன்று வரும். அதாவது சர்வதாரி, பிரபவ போன்ற ஆண்டுகள் ஒருமுறை தொடங்கி முடிந்து மீண்டும் அறுபது ஆண்டுகளுக்குப் பின் மீண்டும் வரும். இதில் ஒரு முறையான ஆண்டுக் கணக்கு இல்லை. இது 60 வருட சுழற்சி என்பதால் இந்த முறை, வரலாற்றுப் பதிவுக்கு உதவுவதில்லை என அறிஞர்கள் கருத்துக் கூறியுள்ளனர்.
- 3) தைப் புத்தாண்டுக்கு கணக்கு உள்ளதா? உண்டு. இது திருவள்ளூர் ஆண்டாக அறிஞர்களால் நிர்ணயிக்கப்பட்டுள்ளது. தமிழகத்தில் 1921-ம் ஆண்டில் மறைமலை அடிகள் தலைமையில் கூடிய அறிஞர் குழு மிக சீரிய ஆய்வுக்குப் பின் திருவள்ளூர் ஆண்டையும் அதன் தொடக்கமாக தை முதல் நாளையும் தெளிவுபடுத்தியது.

வானொலியில் நான் படைத்த நான் மணிக் கடிகை விளக்க நிகழ்ச்சிக்குக் கிடைத்த மாபெரும் வரவேற்பைத் தொடர்ந்து என் எழுத்துக்களை வலைப்பூவில் வைத்திருக்கிறேன். சில பள்ளிகளின், தொடக்கக் கல்லூரிகளின் ஆசிரியர்கள் அதனைப் பார்த்துவிட்டு தங்கள் தமிழ் மாணவர்களுக்கு அதனைப் படிக்கச் சொல்லித் தூண்டியிருப்பது மிகுந்த மன மகிழ்ச்சி தருகிறது. விரைவில் ஆசாரக் கோவை விளக்கமும் தரவிருக்கிறேன். வானொலியில் ஆசாரக் கோவைக்கு நான் அளித்த விளக்கங்களை தினமும் கேட்டு மகிழ்ந்த நேயர்கள் சிலர் அது போல எளிய விளக்கம் இணையத்தில் எங்கு

தேடியும் கிடைக்கவில்லை என்றார்கள். அவர்களின் வேண்டுகோளுக்கேற்ப அதை சேர்க்கவுள்ளேன். இந்த இணைய மாநாடு தொடங்கும் நேரத்தில் அந்த வேலை முடிந்திருக்கலாம்.

எ.கா. 2008-ஆம் ஆண்டுப் பதிவின் ஒரு பகுதி

கற்க கற்க

விளம்பிநாகனார் என்ற புலவர் இயற்றிய நான்மணிக்கடிகை எனும் நூல், 3-ஆம் பாடல் முதல் 106-ஆம் பாடல் வரை வானொலியில் 'கற்க கற்க' நிகழ்ச்சியில் தினம் ஒரு பாடலாக ஒலியேறி வந்தது. 6.8.07 தொடங்கி 26.12.07 வரை நான்மணிக்கடிகையின் ஒவ்வொரு பாடலும் விளக்கத்துடன் ஒலிபரப்பப் பட்டது.

--- நான்மணிக் கடிகை நூலின் நாலாம் பாடல்.

பறைபட வாழா அசுணமா உள்ளம்
குறைபட வாழார் உரவோர் - நிறைவனத்து
நெற்பட்ட கண்ணே வெதிர்சாம் தனக்கொவ்வாச்
சொற்பட வாழாதாஞ் சால்பு

இனி பாடலின் பொருள்:

அசுணம் என்பது மென்மையான இசையை ரசிக்கக்கூடிய ஒருவகைப் பறவை. பறை எனும் கடினமான தாளவாத்தியக் கருவியின் அதிர வைக்கும் சத்தத்தைக் கேட்டால் அந்த இரைச்சல் தாங்காமல் அந்தப் பறவை இறந்து விடுமாம். அதுபோல்தான் சில மனிதர்களும். வாழ்க்கையில் தங்களுக்கு இனிமை மட்டுமே இருக்க வேண்டும் என்று நினைப்பார்கள். மாறாக துன்பம் ஏதேனும் வந்துவிட்டால் உயிரை விட்டு விடுவார்கள். காட்டில் மரங்கள் மிக அடர்த்தியாக வளர்ந்திருந்தால் அங்கு நெல் முளைக்க வழி இல்லை. வளரத் தொடங்கும் போதே அழிந்து விடும். அது போல, ஒருவன் நல்ல செயல்களைச் செய்ய முயலும்போது சுற்றி உள்ளவர்கள் அதனை ஊக்குவிக்காமல் எதிர்மறையாகப் பேசி அவன் மனதைப் புண்படுத்தினால் அவனது முயற்சி ஆரம்பத்திலேயே முடங்கிவிடும்.

தனிப்பட்ட முறையில்:

எனது 'அறிவோமா நாம்' நூலை நானே சொந்தமாக 'துணைவன்' மென்பொருளைக் கொண்டு எழுதி பின்னர் அச்சுக்குக் கொடுக்க முடிந்தது. எனது கடிதங்களை கூகள் துணையுடன் தமிழிலேயே அனுப்புகிறேன். எனது சொற்பொழிவுகளுக்கு Powerpoint presentation எனும் கணினிப் படைப்புகளையே தயாரிக்கிறேன். கணினியை எங்கேயும் தூக்கிச் செல்லத் தேவையில்லை. எல்லாம் ஒரு சிறு thumbdrive கருவியில் அடங்கி விடும் காலம் இது. அது வந்தபோது என்னைப் போன்ற பல தமிழ்ப் பேச்சாளர்களுக்கு மிகுந்த வசதியைத் தந்தது. சிங்கையிலும் மலேசியாவிலும் நூற்றுக்கணக்கான பள்ளிகள், சமூக மன்றங்கள், கோவில்கள் முதலிய இடங்களுக்கு சென்று பல்வேறு தலைப்புகளில் பேசியுள்ளேன். அவற்றின் விவரங்களை இப்போது எனது வலைப்பூவில் கொஞ்சம் கொஞ்சமாக சேர்த்துக் கொண்டிருக்கிறேன்.

எனது சொற்பொழிவுகளுக்கு தகவல் சேகரிக்கவும் இணையம் பேருதவி செய்கிறது. எடுத்துக்காட்டுக்கு சிங்கப்பூரில் அருள்மிகு தண்டாயுதபாணி முருகன் கோவிலில் மண்டலாபிஷேகத்தின்போது நடந்த பட்டிமன்றம் ஒன்றில் இலக்குவன் மகிமை பற்றி நான் பேச வேண்டியிருந்தது. எவ்வளவுதான் நூல்களில் படித்திருந்தாலும் அவசரத்துக்கு வேண்டிய தகவல்களை சேகரிக்க உதவியது, உலகத் தமிழர்கள் பலர் தனித்தனியாக தங்கள் வலைப் பக்கங்களில் தந்திருந்த விவரங்கள்தாம். என் போன்ற நூலாசிரியர்களின் படைப்புகள் காலம் தாண்டி நிலைத்து அடுத்த தலைமுறைகளுக்கும் சென்று சேர உதவுவது இணையம்தான் என்பதைச் சொல்லத் தேவையில்லை. எந்த நேரத்தில் யாருக்கு தேவை என்றாலும் பார்த்து அறிய இணையம் உதவுவதால் அது தமிழ் வளர்ச்சிக்கு பெருந்துணை புரிவதை ஒப்புக் கொள்ள வேண்டும்.

தவிர, நான் பங்கேற்கும் பெரும்பாலான தமிழ் சமூக நிகழ்ச்சிகளும் இணையத்தில் ஒளிபரப்பாகின்றன. அது நூல் வெளியீடாகட்டும், சாலமன் பாப்பையா தலைமையிலான பட்டி மன்றமாகட்டும், கோவில் சொற்பொழிவாகட்டும்; நிகழ்ச்சி முடிந்த கையோடு இணையத்தில் வெளிவருகிறது. திண்ணை போன்ற இணைய பக்கங்களில் சிங்கப்பூர் நிகழ்ச்சிகள் பற்றிய குறிப்புகள், நடந்த நிகழ்ச்சிகள் பற்றிய தகவல், செய்தி, அறிவிப்பு முதலியன விரைவாக வருவதையும் அவற்றை வெளிஇட தமிழ் ஆர்வலர்கள் பலர் ஆர்வத்தோடு முனைவதையும் நாம் அறிவோம்.. Facebook, Twitter, Podcast போன்ற பல வசதிகளையும் பயன்படுத்தும் நம்மவர்கள் அவ்விடங்களில் தமிழையும் கொண்டு வர முனைவது பாராட்டுக்குரியது.

எனது தற்போதைய எனது ஆசிரியப் பணியில்...

கணினியோடு பிறந்து வளரும் இன்றைய இளைய சமுதாயம் தமிழையும் படிக்குமா என்ற கேள்விக்கு விடை எந்த அளவுக்கு நாம் கணினியைப் பயன்படுத்துகிறோம் என்பதைப் பொறுத்துள்ளது என்பதே என் கருத்து. எடுத்துக்காட்டுக்கு என்னுடைய தற்போதைய ஆசிரியப் பணியில் இருவித மாணவர்களை கவனிக்க வேண்டியுள்ளது. இயல்பாக தமிழ் பேசும் கெட்டிக்கார மாணவர்களுக்கு தமிழார்வம் தானே வருகிறது. ஆனால் சற்று திறன் குறைந்தவர்கள் கதை வேறு. ஆங்கிலம், கணிதம் போன்றவற்றுக்கே சிரமப்படும்போது, இரண்டாம் மொழியான தமிழில் அவர்களை ஆர்வம் காட்ட வைப்பது ஒரு சவால்தான். நான் என் மாணவர்களுக்கு கூட வழி தமிழ் எழுத்துகளை அறிமுகம் செய்து வைத்த பொது அவர்களிடையே நல்ல மாற்றத்தைக் காண முடிந்தது. இது கணினி யுகம். கையால் பேனா பிடித்து தமிழ்க் கட்டுரை எழுத சொன்னால் முன்கூட்டி மாணவர்கள், கணினி வழி ஆங்கில விசைப் பலகை கொண்டு தமிழை எழுதும் போது வசதியை உணர்கிறார்கள். அவ்வாறு தமிழை முதன் முதலில் எழுதிய மாணவி சொன்னாள் "This is magic. I will write in Tamil."

மேலும் புதிய கணினி வழிக் கல்வியை இன்று சிங்கப்பூரின் தமிழ் ஆசிரியர்கள் பயன்படுத்தி வருகிறார்கள். Google word document போன்றவற்றை பாடத்திட்டத்தில் பயன்படுத்துகிறோம். எ.கா. கருத்தறிதல் பயிற்சிக்கு, வழக்கமாக பாட நூலில் உள்ள பனுவல்களைப் படித்ததும் கேள்வி பதில்களை சொல்ல வேண்டும், மொழிப் பாடத்தில் பலவீனமான மாணவர்கள் தெரிவு விடைக் கேள்விகளுக்கு (MCQ) பதிலளிப்பர். பெரும்பாலும் ஆர்வம் குன்றிய அவர்களுக்குரிய தெரிவு விடை கேள்விகளை Google word டாகுமென்ட் form முறையில் நானே தயாரித்து தரும்போது சற்று ஆர்வத்தோடு செய்கிறார்கள். தாங்கள் அனுப்பும் பதில் உடனுக்குடன் ஆசிரியரின் கணினியில் பதிவாகி வெண்திரையில் தோன்றும் என்ற எண்ணமே அவர்களை விரைவாக தமிழ்ப்பாடத்தில் ஈடுபடச் செய்கிறது.

மாணவர்கள் Powepoint படைப்புதான் நிறைய செய்கிறார்கள். அதுவே பயனளிக்கவும் செய்கிறது. எ.கா. ராமாயணம் பற்றி எழுது என்றால் ஒருவரும் அக்கறையோடு எழுதுவதில்லை. மாறாக ஒரு 4 பக்கம் கணினியில் ராமாயணம் பற்றிய படைப்பை செய்து அனுப்பு என்றால் உடனடியாக அழகாக செய்கிறார்கள். என் பெரும் நம்பிக்கை என்னவென்றால் காலவோட்டத்தில் தமிழை மறந்த சமுதாயங்கள் திரும்பவும் தமிழைப் பெற இணையம் உதவும் என்பதே. இன்றும் பள்ளியில் தமிழ் இல்லாத அமேரிக்கா போன்ற நாடுகளில் வாழும் தமிழர்கள் தங்கள் பிள்ளைகளிடம் தமிழறிவை ஊட்டவும் பழக்கப்படுத்தவும் இணைய வழி திரைப்படம், பாட்டு போன்றவற்றை பயன்படுத்துகின்றனர் என்றும் தமிழ்ப்பாட மென்பொருள்களைப் பயன்படுத்தி குழந்தைகளுக்குத் தமிழ் கற்பிக்கின்றனர் என்றும் அறியப்படுகிறது. எதிர்காலத்தில் தமிழை வளர்க்கப்போவது கணினி/ இணையம் தான் என்பதில் எந்த சந்தேகமும் இல்லை.

இது போன்ற மாநாடுகளில் நம்மவர்கள் தத்தம் அனுபவத்தைப் பகிர்ந்துகொண்டு மென்மேலும் தமிழின் பெருமையை வளர்க்கவும் தமிழ்ப் புழக்கத்தைப் பரப்பவும் ஆவன செய்வோமாக.

இலங்கையில் மின்னரசாங்கம் நோக்கிய

தமிழ் தகவல் தொழில்நுட்பம்

சவால்களும் சாதனைகளும்

தங்கராஜா தவருபன்

பிரதம நிறைவேற்று அதிகாரி, Speed IT net, Srilanka

ஆசிரியர் – “MyComputer” – Tamil IT Magazine

முகவுரை

தமிழில் கணினி , தமிழ் இணையம் என்று நாம் பேசிவருகின்றோம் இவையெல்லாம் தகவல் தொழில் நுட்பத்தில் தமிழ் மொழியை ஈடுபடுத்தி தமிழ் தகவல் தொழில்நுட்பத்தின் ஊடாக நாம் எமது நடவடிக்கைகளை செய்வதற்கே என்பது யாவரும் அறிந்தது

இன்றைய தகவல்தொழில்நுட்ப உலகில் மின்னரசாங்கம் என்ற பதம் மிகப்பிரபல்யமாகி வருகின்றது. இந்நிலையில் உலகமெங்கும் நாடுகள் தமது அரசு நடவடிக்கைகளை மின்னரசு நடவடிக்கைகளாக மாற்றிவருகின்றன. இந்தவகையில் இலங்கையும் தனது அரசு நடவடிக்கைகளை மின்னரசு நடவடிக்கைகளாக மாற்றுவதற்கான முயற்சிகளில் நீண்டகாலமாக ஈடுபட்டுவருகின்றது. ஈழத்தில் மின்னரசுக்கான ஏது நிலைகள் என்ற தலைப்பில் 2004 மாநாட்டில் என்னால் ஒரு கட்டுரை வழங்கப்பட்டது அதிலே ஆரம்ப கட்ட நடவடிக்கைகள் அன்றைய நிலை குறித்து விபரித்திருந்தேன்

தற்போது இலங்கையில் மின்னரசு நடவடிக்கை முயற்சிகளில் பல்வேறு முன்முயற்சிகள் எடுக்கப்பட்டு சாதிக்கப்பட்டிருக்கின்றன. இவ்வேளையில் மின்னரசுக்கான மின்சமூகத்தினை தயார்படுத்தவேண்டிய தேவை இங்கு எழுந்திருக்கின்றது. முன்னரே இந்த தேவை இருந்து அதற்கான முயற்சிகள் பலதரப்பினராலும் எடுக்கப்பட்டு வந்திருந்தாலும் இந்த மின்சமூகத்தினை கட்டியெழுப்புவது தற்போது போரில் இருந்து மீண்டிருக்கும் இலங்கைக்கு மிகப்பெரிய வேலைத்திட்டம் என்றே சொல்ல முடியும்

அந்தவகையில் மின்னரசாங்கத்தினை நோக்கிய தமிழ் தகவல்கள் தொழில்நுட்பத்துறையில் எமது மக்களும் அரசு தனியார் நிறுவனங்களும் எவ்வாறான சவால்களை சந்தித்தனர் சந்திக்கின்றனர் சந்திக்கப்போகின்றனர் என்பதனையும் இதுவரை சாதித்தவை என்ன என்பதையும் ஆய்வுசெய்வதாக இக்கட்டுரை அமைகிறது.

இலங்கையில் மின்னரசாங்கம்

தகவல் தொடர்பாடல் தொழில்நுட்பத்தினைப் பயன்படுத்தி அரசாங்கம் ஒன்றின் நடவடிக்கைகளை ஒருங்கமைத்தலும் பராமரித்தலும் மின்னரசாங்கம் எனக் கருதப்படும்.

இலங்கையில் மின்னரசாங்கம் இருக்கின்றதா அல்லது கட்டமைந்து வருகின்றதா என்று கேட்டால் கட்டமைந்து வருகின்றது என்றே கூறமுடியும். இலங்கை அரசு மின்னரசு நோக்கிய தனது பயணத்தின் அடிப்படையான மைல் கற்களைத் தாண்டி விட்டது எனக் கூறலாம். முக்கிய சான்றாக மின்னரசுக்கான கொள்கைகளும் நடைமுறைகளும் (E-government Policy) இலங்கை அமைச்சரவையினால் டிசம்பர் 2009 இல் நிறைவேற்றப்பட்டிருப்பதை கூறலாம்.

இலங்கைத் தகவல் தொடர்பாடல் தொழில்நுட்ப முகவரான்மை நிலையம் (ICTA)

மின்னரசு நோக்கிய முதலாவது நகர்வு ஆக 1983இல் நிறைவேற்றப்பட்ட தேசிய கணினிக் கொள்கையினக் கூறலாம். இதன் மூலம் இலங்கை அரசு தகவல் தொடர்பாடல் தொழில்நுட்பத்தினது தேவையினை அங்கீகரித்தது. தொடர்ச்சியான நடவடிக்கையின் பயனாக 2003 ம் ஆண்டின் 27ஆம் இலக்க தகவல் தொடர்பாடல் மற்றும் தொழில்நுட்ப சட்டத்தின் கீழ் இலங்கை அரசாங்கத்தினதும் உலகத்தினதும் ஆதரவு நிதியுதவிகளுடன் இலங்கைத்தகவல் தொடர்பாடல் தொழில்நுட்ப முகவரான்மை நிலையம் (Information Communication Technology Agency) ஆரம்பிக்கப்பட்டு நாட்டின் தகவல் தொழில்நுட்பம் சம்பந்தமான செந்தரக்(Standard) கொள்கைகள் வகுக்கப்பட்டு செயற்றிட்டங்கள் முன்னெடுக்கப்படுகின்றன.

இந்த நிறுவனமே 1984 ம் ஆண்டு 10ம் இலக்க பாராளுமன்ற சட்டமொன்றினால் உருவாக்கப்பட்ட தகவல் தொழில்நுட்ப மன்றத்தின் சட்ட பூர்வமான அடுத்த நிறுவனமாக அடையாளம் காணப்பட்டுள்ளதோடு சனாதிபதி செயலகத்தின் நேரடிக்கண்காணிப்பின் கீழ் தொழிற்படுகின்ற இலங்கை அரசு உச்ச நிலை நிறுவனமாக தொழிற்படுகின்றது.

2008ம் ஆண்டின் 33ம் இலக்க தகவல் தொடர்பாடல் தொழில்நுட்ப திருத்தச்சட்டத்திற்கு அமை வாக இந்நிறுவனமே தேசிய தகவல் தொடர்பாடல் சட்டத்தினை வகுப்பதிற்கு உள்ளக அமைச்சு மட்ட குழுவுக்கு சிபார்சுகள் வழங்க அதிகாரம் அளிக்கப்பட்டிருக்கின்றது.

மின்னரசில் தமிழ் தகவல் தொழில்நுட்பம்.

இவற்றிலிருந்து ஒன்றைத் தெளிவாகத் தெரிந்து கொள்ள முடிகின்றது. அதாவது இலங்கையின் மின்னரசு நோக்கிய தமிழ் தகவல் தொழில்நுட்பத்திலும் இவ்வமைப்பு முக்கிய வகிபாகத்தினைக் கொண்டிருக்கின்றது.

இலங்கையின் ஆட்சிமொழிகளில் தமிழும் இடம்பெற்றிருப்பதனால் தமிழ் மொழிக்கு சகல நிலைகளிலும் இடம் வழங்கப்பட்டிருக்க வேண்டிய தேவை உள்ளது. இந்நிலையில் மின்னரசாங்க செயற்பாடுகளிலும் தமிழ் பயன்படுத்தப்படவேண்டியுள்ளது. இதன் காரணமாக தமிழ்த் தகவல் தொழில்நுட்பத்தின் உலகளவிய செந்தரங்களையும் மாற்றங்களையும் உள்வாங்கி தனது பிரசைகளின் எதிர் பார்ப்புக் களையும் திருப்தி செய்து சிறந்த மின்னரசினைக் கட்டியெழுப்புவதில் இலங்கை அரசு ஈடுபட்டிருக்கின்றது.

இலங்கை அரசின் செந்தரங்கள்(Standards) மற்றும் பயன்பாடு.

எழுத்துரு மற்றும் விசைப்பலகை

தற்போதைய நிலையில் செந்தரங்களினைப் பொறுத்தவரையில் தமிழ் தகவல் தொழில்நுட்பம் தொடர்பில் 2008ம் ஆண்டு **SLS 1326** நிர்ணயத்திற்கு அமைவாக அனைத்து அரசு நிறுவனங் களும் கட்டுப்பட வேண்டிய தேவை உள்ளது. இதற்கு அமைய எழுத்துரு (குழுNu) பாவனைக்கு யுனிகோட் (Unicode) செந்தரத்தினை அரசு ஏற்றிருக்கிறது. விசைப்பலகையினை (KeyBoard) பொறுத்தவரை முன்னர் பாவனையில் இருந்த ரங்கநாதன் விசைப்பலகை(பாமினி எழுத்துரு விசைப்பலகை) முறையில் சுமாராக பத்திற்கும் குறையாத மாற்றங்களைக் கொண்ட புதிய விசைப்பலகை முறையினை அரசு பரிந்துரை செய்திருக்கின்றது.

2004ஆம் ஆண்டில் ICTA யின் பரிந்துரையில் தமிழ்99 (Tamil99) விசைப்பலகை முறை ஏற்றுக் கொள்ளப்பட்டிருந்தது எனினும் மக்களிடையே இதன் செல்வாக்கு சரிந்ததன் காரணமாக அல்லது இது ஏற்றுக்கொள்ளப்படாததன் காரணமாகவும் இது பற்றி ஆராய அமைக்கப்பட்ட குழுவின் சிபார்சின் அடிப்படையிலும் மட்டுப்படுத்தப்பட்ட அளவில் மேற்கொள்ளப்பட்ட மீளாய்வின் அடிப்படையிலும் புதியவிசைப்பலகை முறை பரிந்துரை செய்யப்பட்டிருக்கின்றது.

புதிய மாற்றங்களை மக்கள் உள்வாங்குவது அவ்வளவு கடினமாக இருக்காது. ஏனெனில் ஏற்கனவே பாமினி எழுத்துரு விசைப்பலகை முறையினையே இலங்கைத் தமிழர் பயன்படுத்தி வந்தனர். ஏன் இன்னும் பயன்படுத்தி வருகின்றனர் என்றே சொல்ல முடியும். ஆவணங்களிலும் தொழில்நுட்ப வியலாளர்கள் மத்தியிலும் இம்முறை ரங்கநாதன் விசைப்பலகை முறை எனச் சொல்லப்பட்டாலும் இதன் தோற்றுவாய் குறித்தோ இந்த விசைப்பலகை முறை பற்றியோ மேலும் நான் ஆராய விரும்பவில்லை. ஏனெனில் இக்கட்டுரை மின்னரசு குறித்தது. மேற்படி ௭௫௭ 1326 நிர்ணயத்திற்கு அமைய இலங்கையில் எழுத்துரு, விசைப்பலகை முறைகளுக்கு முற்றுப்புள்ளி வைக்கப்பட்டிருக்கிறது.

கலைச்சொல்லகராதி

ஒரு மின்னரசு நடவடிக்கையில் முக்கிய பங்கு வகிக்கும் அடுத்த விடயம் கலைச்சொல்லகராதி. கணினியில் தமிழ்மொழி பயன்பாட்டிற்கு துணைபுரியப்போவது இதுவேயாகும். தமிழ் ஒரு வில்லங்கமான மொழி. தமிழ் மொழிபெயர்ப்பு என்பது வெறுமனே மற்ற மொழியிலிருந்து நேரடியாக மொழிபெயர்ப்பதை நோக்காகக் கொண்டிருக்க முடியாது. மொழிபெயர்ப்பானது தற்காலத்திற்கு இயைபானதாகவும் கருத்துப் பொழிவுள்ளதாகவும் இருக்க வேண்டும்.

பல்வேறு கலைச்சொல்லகராதி உருவாக்க முயற்சிகள் உலகளாவிய ரீதியில் முன்னெடுக்கப்பட்டன. தமிழ் மொழிவழக்கில் இரண்டு முக்கிய பிரிவுகள் காணப்படுகின்றன. ஒன்று இலங்கை வழக்கு (ஈழத்தமிழர் வழக்கு) மற்றையது இந்தியத் தமிழர்; வழக்கு. இந்த இழுபறிக்கு மத்தியில் இந்திய இலங்கை தமிழ் அறிஞர்களினதும் தகவல் தொழில்நுட்பவியலாளர்கள் சிலரினதும் கூட்டு முயற்சியில் உருவான தமிழ் தகவல் தொழில்நுட்ப கலைச்சொல் அகரமுதலியினை இலங்கை அரசு கரும மொழிகள் ஆணைக்குழு புத்தகமாக 2000 ஆண்டில் வெளியிட்டது.

இதனைவிட யாழ் பல்கலைக்கழகத்தினால் உருவாக்கப்பட்ட கலைச்சொல் அகரமுதலி ஒன்றும் அறியக்கிடக்கிறது. தனிப்பட்ட முயற்சிகளும் முன்னகர்த்தப்படுகின்றன. LAKAPPS இனாலும் சில முயற்சிகள் எடுக்கப்படுகிறது

அரசினால் 2000 இல் வெளியிடப்பட்ட அகராதியே பெருமளவில் பாவனையில் உள்ளது. இதனையே மொழிபெயர்ப்பாளர்கள் கருத்தில் கொள்கின்றனர். இதனை விடவும் தனிநபர் முயற்சிகளினாலும் இணையத்தளங்களில் சிதறிக்கிடக்கும் முயற்சிகளில் இருந்தும் சிலர் கலைச்சொற்களைப் பயன்படுத்துகின்றனர்.

மேற்சொன்ன விசைப்பலகை, எழுத்துரு, கலைச்சொல்லகராதி போன்ற விடயங்களில் பரிந்துரைகள் ஒருபுறம் இருக்க பாவனை, நடைமுறைப்படுத்தல் ஆகிய விடயங்களில் என்ன நிலவரம் என நோக்க வேண்டியிருக்கிறது.

மின்சமூகத்தில் தமிழர், மின்னரசில் தமிழின் நிலை

மின்னரசு யாருக்காக நிறுவப்படுகின்றது. நாட்டு மக்களுக்காகவல்லவா? அந்நாட்டு மக்கள் மின்னரசின் ஊடாக தமது நடவடிக்கைகளை மேற்கொள்ளும் போது மின்சமூகத்தினுள் நுழைகின்றார்கள். ஆக இந்த மின் சமூகம் மின்னரசு நடவடிக்கைகளை கையாள வல்லவர்களாக இருக்க வேண்டும். முதலில் மின் சமூகம் கணினி அறிவுடையவர்களாக இருக்க வேண்டும். அந்த சமூகம் சார்ந்தவர்கள் தமது தாய் மொழியினை ஊடகமாக கணினியில் பயன்படுத்தக் கூடியவர்களாக மாற வேண்டும். அல்லது மின்னரசு மொழிகளில் ஏதாவதொன்றைப் பயன்படுத்தக் கூடியவர்களாக இருக்க வேண்டும்.

இங்கு மின்சமூகம் என்னும் போது வெறுமனே பயனாளர்களை மட்டும் குறிக்காது. அரசு இயந்திரத்தி லுள்ள அலுவலர்களினையும் குறிக்கும். அது சாதாரண அரசு ஊழியர் தொடங்கி அதியுயர் பீடம் வரையும் குறிக்கும். இந்நிலையில் நான் தகவல் தொழில்நுட்பம் பற்றியே தலைப்பில் விளித் திருப்பதனால் தமிழ் மக்களது தகவல் தொழில்நுட்ப பயன்பாடு மற்றும் அறிவு பற்றியே கவனம் செலுத்த வேண்டியவனாயிருக்கிறேன்.

ஆரம்பத்தில் இருந்தே தமிழர்கள் தமது அரசியல் உரிமைகளுக்கான போராடி வந்த போதிலும் தகவல் தொழில் நுட்ப அறிவை வளர்த்துக் கொள்வதில் பின்னிற்கவில்லை. இருந்த பொழுதிலும் இத் தொழில்நுட்பம் சகல மட்டத்தினரையும் அடைவதில் தான் பிரச்சினைகள் இருந்தன. எது எப்படியிருந்த பொழுதிலும் தமிழ் தகவல் தொழில்நுட்பத்தில் ஈழத்தமிழர்கள் குறிப்பாக புலம்பெயர் தமிழர்கள் தமது பங்கினைக் காலத்திற்குக் காலம் குறிப்பிடத்தக்க அளவில் வழங்கி வந்துள்ளனர் என்பதனை தமிழ்த் தகவல் தொழில்நுட்ப வரலாற்றினை பின்னோக்கி பார்ப்பதன் மூலம் அறிந்து கொள்ள முடியும்.

இந்நிலையில் இலங்கையில் மின்னரசினை எடுத்து நோக்கில் இலங்கை மின்சமூகத்தில் தமிழர்கள் உள்வாங்கப்பட்டு விட்டார்களா? மின்னரசில் தமிழுக்குரிய இடம் கிடைத்து விட்டதா? என்ற வினாக்களுக்கு விடை தேட முற்படும் போது, மின்சமூகத்தில் அனைத்து தமிழ் சமூகமும் உள்வாங்கப்படவில்லை என்றே பதில் வழங்க முடியும். இந்நிலை தமிழர்களுக்கு மட்டுமல்ல சிங்களவர்களுக்கும் கூட பொருந்தும். இருப்பினும் பாதிப்பு தமிழர்களுக்கே அதிகம். தமிழர்கள் மேலே சொன்னவாறு உரிமைப்போராட்டத்தில் சிக்கித் தவித்த நிலையில் அபிவிருத்தித் தொழில்நுட்பங்கள் வெறும் பேச்சளவிலேயே இருந்தது வந்திருக்கின்றது. அவர்களால் சகல விடயங்களையும் மற்றவர்கள் போல நுகர முடியாத நிலை இருந்தது.

மின்னரசில் உரிய இடம் தமிழுக்குக் கிடைத்ததா என்றால் கிடைத்தது ஆனால் அதனை அவர்கள் பயன்படுத்த முடியவில்லை என்று தான் கூறமுடியும். அதற்கு முக்கிய காரணம் போர்தான் என மீளவும் தெளிவுபடுத்த வேண்டிய அவசியமில்லை. இதன் காரணமாக தமிழ் தகவல் தொழில்நுட்ப முயற்சிகளை முன்னெடுப்பதிலும் மின்னரசில் தமிழிற்கான இடத்தினை உறுதி செய்வதற்குமுரிய ஆளணியினர் மிகச் சொற்பமாகவே இலங்கையில் காணப்பட்டனர். இருந்த பொழுதிலும் மற்ற மொழிகளுடன் போட்டி போடக்கூடிய அளவிற்கு தமிழ்த் தகவல் தொழில்நுட்பம் இருக்கின்றது. அதற்கு குறிப்பிட்ட சொற்ப ஆளணியினர் குறிப்பிடத்தக்களவு பங்கினை வகிக்கனர் என்பதை ஏற்றுக்கொள்ள வேண்டியிருக்கிறது.

எம் முன் இருக்கும் சவால்களும்; தீர்வுகளும்;

1. சகலருக்கும் தொழில்நுட்ப அறிவை வழங்குதல்
2. மின்னரசில் தமிழுக்குரிய இடத்தினை உறுதிப்படுத்துதல்.
3. மின்னரசினை மேம்பட்ட நிலைக்கு உயர்த்துதல்.

போரின் வடுக்களிலிருந்து மீண்டு கொண்டிருக்கும் மக்கள் படிப்படியாக தமிழ் தகவல் தொழில்நுட்பத் தினுள் உள்வாங்கப்பட வேண்டிய நிலையில் அரசு பல திட்டங்களை கடந்த காலங்களில் முன்னெடுத்திருக்கின்றமையினை நாம் மறுக்க மடியாது.

சகலருக்கும் தகவல் தொழில்நுட்ப அறிவையும் கணினியுடன் இணைந்த ஆங்கில் அறிவையும் வழங்கும் முகமாக 2009ம் ஆண்டினை தகவல் தொழில்நுட்ப ஆங்கில் ஆண்டாக பிரகடனப் படுத்தி செயற்பாடுகளைச் செய்தது ஆயினும் தூரதிஸ்டவசமாக போர் உச்சமடைந்திருந்ததன் காரணமாக வடக்குக் கிழக்கு தமிழர் பிரதேசங்களில் இச் செயற்றிட்டங்கள் ஓரளவுக்கே வெற்றியளித்தது. தற்போது கிழக்கின் உதயம், வடக்கின் வசந்தம் என திட்டங்கள் முன்னெடுக்கப்படுகின்ற நிலையில் இந்தத் திட்டங்களின் ஒரு விளைவாக தமிழ் மக்கள் அனைவரும் தகவல் தொழில் நுட்ப அறிவை பெற்றுக் கொள்வதற்கான ஏது நிலைகள் தோற்றுவிக்கப்பட்டிருக்கின்றன.

ICTA இதற்கான வேலைத்திட்டங்களை ஆரம்பித்திருக்கின்றது என்பதை அண்மைக்கால நடவடிக்கைகள் கட்டியம் கூறுகின்றன. இது மேலும் முனைப்பு பெற வேண்டும் என்பது அனைவரதும் அவா. பெயரளவில் தொடங்கப்பட்டு எமது பிரதேசங்களில் சரிவர நடைமுறைப் படுத்தப்படாதவைகள் மீண்டும் முனைப்புடன் செயற்படுத்துவதற்கான காலம் தற்போது கனிந்திருக்கின்றது.

தமிழிற்கான இடம்.

ஒவ்வொரு மின்னரசின் நடவடிக்கையின் போதும் போர்ப்பிரதேசங்கள், அரசு கட்டுப்பாடற்ற பிரதேசங்கள் என தட்டிக்கழிக்கப்பட்ட நிலை இனி இருக்காது. அதே போல எதிர்காலத்தில் நிகழக்கூடிய தமிழ் எழுத்துருக்கள், விசைப்பலகை, ஆகிய செந்தரங்களை மீளாய்வு செய்யும் நடவடிக்கைகளிலும் கலைச்சொல்லகராதி நடவடிக்கைகளிலும் குறிப்பாக வடக்குக் கிழக்கு பிராந்திய மக்களின் பங்கு திருப்திகரமான வகையில் உறுதி செய்யப்பட வேண்டியிருக்கும். அவர்களிடத்தே இருக்கும் பயன்பாட்டு முறைகள் வழக்கு முறைகளை உரிய முறையில் உள்வாங்கி மின்னரசு செயற்பாடுகள் முன்னெடுத்துச் செல்லப்பட வேண்டியுள்ளது.

தற்போது பரிந்துரை செய்யப்பட்டுள்ள செந்தரங்கள் மற்றும் கலைச்சொல்லாக்கங்கள் வடக்குக் கிழக்கில் முழுமையாக பின்பற்றப்படாதிருக்கின்றன அல்லது அறியப்படாமல் இருக்கின்றது. யுனிக்கோட், எழுத்துருப்பாவனை ஏற்றுக்கொள்ளப்பட்டிருக்கின்ற போதிலும் விசைப்பலகையினைப் பொறுத்தபலர் பழைய பாமினி எழுத்துரு விசைப்பலகையினையே பயன்படுத்துகின்றனர். ஒருசிலர் தமிழ் 99 விசைப்பலகையினையும் பயன்படுத்துகின்றனர். அதே போல் கணினி என்ற கலைச்சொல் ஆக்கத்தினை மறுத்து கணினி என்ற பதத்தினையே பலர் பயன்படுத்தி வருகின்றனர். அதற்காக மாற்றங்களை முற்றாக ஏற்றுக்கொள்ளவில்லை என்றோ அறியப்படவில்லை என்றோ கூறமுடியாது. சிலர் மாறிவிட்டார்கள். பலர் மாறவில்லை என்று தான் கூறமுடியும்.

மக்கள் மனங்களில் மாற்றங்களை ஏற்படுத்தவும் ஒன்றுபட்ட செந்தரத்தினை பின்பற்றச் செய்வதற்கும் காலத்திற்குக் காலம் சில சஞ்சிகைகள், கணினி நிறுவனங்கள் ஏன் அரசும் பாடுபட்டிருக்கின்றன என்னும் இது திருப்திகரமான இருக்கவில்லை என்பதே வெளிப்படை.

எனவே கணினி நிறுவனங்கள், சஞ்சிகைகள், பத்திரிகைகள், இலத்திரனியல் ஊடகங்கள், பல்கலைக் கழகங்கள், கல்விச்சாலைகள், இது விடயத்தில் ஒன்றுபட்டு செயற்பட வேண்டியிருக்கின்றது. அரசாங்கம் ICTA ஊடாக இது தொடர்பான செயற்றிட்டங்களை முன்னெடுக்க வேண்டியிருக்கின்றது.

மின்னரசாங்கத்தினைப் பொறுத்தவரை ஒரு பொதுப்பிரச்சினையும் உள்ளது. செயற்றிட்டங்களைப் பொறுப்பெடுக்கும் தனியார் நிறுவனங்கள் மும்மொழியில் தேர்ச்சி பெற்றவர்களை கொண்டிருக்க தவறியுள்ளமை தான் அது. இதனால் இவர்களது மொழிபெயர்ப்புக்கள் தமிழுக்கு பெரும் பாதிப்பை ஏற்படுத்தி விடுகின்றது.

உதாரணமாக எடுத்துக்கொண்டால் ஒரு மென்பொருள் அல்லது இணையத்தளம் ஆனது தமிழ், ஆங்கிலம், சிங்களம் என மூன்று தெரிவுகளைக் கொண்டிருந்ததால் அவற்றின் குறித்த மொழி ஊடான செயற்பாட்டுத் தன்மை கேள்விக்குறியாக உள்ளது. தமிழ் வேலை செய்யும் சிங்களம் வேலை செய்யாது அல்லது ஆங்கிலம் தவிர்ந்த ஏனையவை வேலை செய்யாது. இதற்கு சிங்களம் விதிவிலக்கல்ல. தமிழ் தகவல் தொழில்நுட்ப பயன்பாட்டுக்கு உள்ளதை போன்ற நிலையே அதற்கும் உள்ளதை மறுக்க முடியாது. இதற்கு அரசு நிறுவனங்கள் குறித்த மொழி தொடர்பில் வழங்கும் தகவல் பற்றாக்குறை மற்றும் சிறந்த மொழிபெயர்ப்பாளர்கள் இன்மையும் காரணங்களாகக் கொள்ளப்படலாம்.

மின்னரசினை மேம்பட்ட நிலைக்கு உயர்த்துதல் பரவலாக்கல் -

தற்போதுள்ள மின்னரசுக் கட்டமைப்புகளில் மாற்றம் கொண்டுவரப்பட வேண்டும். அவை சகல துறைகளுக்கும் பரவலாக்கப்பட வேண்டும். அதேவேளை மின்னரசில் பங்கெடுக்கும் மின் சமூகத்தினை கட்டியெழுப்பப்பட வேண்டும். இவையே மின்னரசினை மேம்பட்ட நிலைக்கு உயர்த்துவதற்கான வழிகளாகும்.

இதற்கான கட்டியம் தான் அமைச்சரவையினால் அங்கீகரிக்கப்பட்ட மின்னரசுக்கொள்கைகள் என கூற முடியும்.

மின்னரசு நடவடிக்கைகளில் குறிப்பிடுபம்படியாக தற்போதும் செயற்படுநிலையில் இருக்கின்ற சில செயற்றிட்டங்களையும் அவை பற்றிய விளக்கங்களையும் கீழே நோக்குவோம். அவை சரிவரச் செயற்படுத்தப்படுமாயின் இலங்கையில் மின்னரசு விரைவில் உன்னத நிலையினை அடையும்.

- (1) 2003-2005 காலப்பகுதியில் உருவாக்கப்பட்ட **e-Srilanka** அபிவிருத்தி கருத்திட்டம் ICTA ஊடாக செயற்படுத்தப்பட்ட வருகின்றது. இது சமூகத்தில் அனைத்து தரப்பினருக்கும் தகவல் தொழில்நுட்பத்தின் நன்மைகளை வழங்குவதோடு பன்னாட்டு நன்கொடை யாளர்களின் நிதி உதவிகளின் மூலம் செயற்படுத்தப்படும் திட்டங்களின் ஊடாக தேவையான உட்கட்டமைப்பினை உருவாக்கிக் கொள்வதற்கும் மின்னரசாங்க சேவைகள் ஆரம்பிக்கப்படுவதற்கான புறநிலைகளை உருவாக்கும். (இணைப்பு 01)
- (2) பாடசாலைகளில் தகவல் தொடர்பாடல் தொழில்நுட்பம் ஒரு பாடமாகப் புகுத்தப்பட்டு தரம் 10, 11 மற்றும் உயர்தர மாணவர்களுக்கும் கணினிக்கல்வி வழங்கப்படுகின்றது. இது முறையே ICT, GIT என பெயரிடப்பட்டிருக்கின்றது. மேலும் பல்கலைக்கழக தெரிவிற்கான பாடநெறிகளில் கூட தகவல் தொழில்நுட்பம் ஒரு பாடமாகக் கொள்ளப்பட ஏற்பாடாகியுள்ளது. இதனைவிடவும் **Schoolnet** எனும் திட்டத்தின் ஊடாக சகல பாடசாலைகளும் இணைய வலைப் பின்னலில் இணைக்கும் செயற்பாடுகள் முன்னெடுக்கப்படுகின்றன. (இணைப்பு 02,03,04,05)
- (3) மின் சமூகத்தின் முதுகெலும்புகளான மாணவர்களின் கல்விக்கு முதலிடம் அளிக்கப்பட்டுள்ளது. இணையவழி தொலைக்கல்வியினை NODES,DEMP திட்டங்களின் ஊடாக வழங்குகிறது அதனை கல்வியமைச்சு வழிநடத்துகிறது. இதில் ஒரு புரட்சிகர அம்சமாக இலங்கையில் முதன் முதலாக தமிழ் மூல இணையவழி பட்டக்கல்வியினை அறிமுகப்படுத்தி யுள்ளனர்.யாழ்.பல் கலைக்கழகத்தின் முகாமைத்துவ கற்கைகள் வணிக பீடத்தினரின் “வியாபார முகாமைத்துவ மாணி” பட்டக் கற்கைநெறி இதுவாகும். (இணைப்பு 06,07)
- (4) எமது மொழிகள்(emathumozihal) என்ற இணையத்தளம் ஊடாக தமிழ் எழுத்துரு மற்றும் விசைப்பலகை முறை பாவனை சம்பந்தமான தகவல்கள் மக்களுக்கு வழங்கப்படுகின்றது. (இணைப்பு 08)
- (5) மென்பொருட்கள் மற்றும் செயலிகளை உள்ளூர் மயமாக்கும் பொருட்டு “LAKAPPS” செயற் திட்டம் முன்னெடுக்கப்படுகிறது இவர்கள் இது தொடர்பான பயிற்சிகளையும் பயனர்களுக்கு வழங்கிவருகின்றனர். (இணைப்பு 09)
- (6) LK Domain Registry (இலங்கை ஆள்கள பெயர் பதிவு மையம்) தமிழ் மொழியிலான IDN முயற்சிகளில் ஈடுபட்டு ICANN அமைப்பில் இதற்கான அங்கீகாரத்தினையும் பெற்றுவிட்டது. உலகின் முதலாவது தமிழிலான நாடுகளின் உயர்நிலை ஆள்கள பெயரினை உருவாக்கி யுள்ளது. இனிமேல் “தளம்.அரசு.இலங்கை” என அழைத்தால் அரசாங்க தளம் தோன்றும். “.இலங்கை” என்பதே இலங்கை நாட்டுக்கான தமிழ் IDN ஆகும். “தமிழ்dns.இலங்கை” அல்லது www.idns.lk இல் இது பற்றி மேலதிக விபரத்தினை பார்வையிடலாம். (இணைப்பு 10,11)
- (7) நெனசல (NENASALA) திட்டத்தின் ஊடாக இலங்கையின் கிராமங்களை இணைக்கும் திட்டம் நடைமுறைப்படுத்தப்படுகிறது. (இணைப்பு 12)
- (8) விதாதா வள நிலையங்கள் ஊடாக கிராமத்துக்கு தொழில் நுட்பம் என்ற தொனிப் பொருளில்; கிராம மக்களுக்கு தகவல் தொழில்நுட்ப கல்வி உள்ளிட்ட பயிற்சிகள் வழங்கப் படுகின்றது. (இணைப்பு 13)
- (9) இலங்கை அரசாங்க தகவல் நிலையத்தின்(GIS) ஊடாக இலங்கையின் சகலமொழி மக்களுக்கும் சகலவிதமான தகவல்களும் வழங்கப்படுகின்றன. மின்னரசு நடவடிக்கையில் இது மிகப்பெரிய செயற்பாடாக கொள்ளலாம். (இணைப்பு 14)

இவைதவிர அரசாங்கத்தின் உத்தியோக பூர்வ இணையத்தளத்தின் ஊடாக சகல திணைக்களங்களும் அமைச்சுக்களும் ஒருங்கிணைக்கப்படுகின்றன. (இணைப்பு 15)

(10) அரசு வர்த்தமானி அறிவித்தல்கள் மும்மொழியிலும் அரசு அச்சுக் கூட்டுத்தாபன இணையத்தளத்தின் ஊடாக இலத்திரனியல் வடிவில் பகிரப்படுகின்றன. (இணைப்பு 16)

(11) மேலும் சமீபத்தில் உருவாக்கப்பட்ட HAPPYLIFE இணையத்தளத்தின் ஊடாக சுகாதார விழிப்புணர்வு தகவல்கள் நாட்டின் பிரசைகளுக்கு மும்மொழிகளும் பகிரப்படுகின்றன. (இணைப்பு 17)

மேற்சொன்ன திட்டங்களில் ஒரு சில தவிர்ந்தவை முற்றுமுழுதாக தமிழிற்கோ சிங்களத் திற்கோ இடமளிக்கத்தவறியிருக்கின்றன பல பகுதியாகவே மொழி பிரயோகத்திற்கு ஆட்பட்டிருக்கின்றன எனவே இவை நடுத்தர வர்க்க மக்களுக்கே பெரிதும் உதவிகரமானதாக இருக்கும் சமூகத்தின் அடித்தட்டு வர்க்கத்தினரையும் இத்திட்டங்கள் சென்றடைய வேண்டுமாயின் தமிழ் தகவல் தொழில் நுட்பம் மற்றும் சிங்கள தகவல் தொழில்நுட்பம் முழுமையாக அமுல்படுத்தப்பட வேண்டும்.

குறிப்பாக பரீட்சைத்திணைக்களம், தேர்தல் திணைக்களம், கல்வித்திணைக்களம் உள்ளிட்டவை தமிழ் மொழியிலான இணையத்தள சேவைகளையும் வழங்க முன்வரவேண்டும்.

வெறுமனே இணையத்தளங்களினூடாக மாத்திரம் மின்னரசு கட்டியெழுப்பப்பட முடியாது அரசு இயந்திரத்தில் ஆதி முதல் அந்தம் வரை தகவல் தொடர்பாடல் தொழில்நுட்பம் ஆதிக்கம் செலுத்த வேண்டும். அலுவலர்கள் சகலரும் பயிற்சியுடையவர்களாக மாற்றப்பட வேண்டும் அந்த வேளையில் மின்சமூகத்திணையம் கட்டியெழுப்ப வேண்டும் இன்னமும் கணினி மயப்படுத்தப்படாத அரசு செயற்பாடுகள் இருக்கவே செய்கின்றன.

கண்காட்சிகளினை நகரங்களில் ஏற்பாடு செய்வதன் மூலம் மட்டும் மின்னரசுக்கான மின் சமூகத்திணை கட்டியெழுப்ப முடியாது கிராமம் தோறும் மின்னரசுக்கான அடித்தளம் அமைக்கப்படவேண்டும் திணைக்களம் தோறும் மின்னரசுக்கான அறைகூவல் விடுக்கப்பட வேண்டும். அதிகாரிகளின் மனங்களும் மாற வேண்டும் அரசு மட்டும் இயங்கிப்பயனில்லை அனைவரும் ஒன்றுபடவேண்டும்.

மேற்சொன்ன வகையில் பல செயற்றிட்டங்கள் மின்னரசுக்கான மேம்பாட்டை மையங்கொண்டு ஆரம்பிக்கப்பட்டுள்ளதை அவதானிக்கக்கூடியதாக உள்ளது. இருக்கின்ற இந்த செயற்றிட்டங்களை; சகல இடங்களிலும் திறம்பட செயற்படுத்துவதன் மூலமும் சகலரையும் அரவணைத்து செல்வதன் மூலமும் மின்னரசு முயற்சிகளும் மின்னரசுக்கான மின் சமூகத்தை வளப்பதற்கான முயற்சிகளும் ஓரிடத்தில் சந்தித்து மின்னரசாங்கத்தில் ஆற்றல் உள்ள மின்சமூகம் கால்பதிக்கும் நிலையினை ஏற்படுத்தலாம். இதற்கான பொறுப்பு நாட்டுமக்கள் சகலருக்கும் உண்டு.

உசாத்துணைகள்

1. e-Government: The Singapore- Arun Mahizhnan & Narayanan Andiappan Tamil Internet 2002, California, USA
2. அரசாங்கத் துறையில்; தகவல், தொடர்பாடல் தொழில்நுட்ப பயன்பாட்டுக்கான கொள்கையும் நடைமுறைகளும் –ICTA ,December 2009
3. E-governance: Tackling the Hurdles, N Jeyaratha & CK Santhakumar Tamil internet 2003, Chennai
4. Tamil Localisation Process – A case study, Kengatharaiyer Sarveswaran & Gihan Dias Tamil Internet 2009

இணைய இணைப்புகள் (30.04.2010 அணுகப்பட்டது)

- 01: <http://www.icta.lk>
- 02: <http://www.nie.lk>
- 03: <http://www.nie.sch.lk/ebook/s11tim33.pdf>
- 04: <http://www.nie.sch.lk/ebook/s12syl41.pdf>
- 05: <http://www.Scoolnet.lk>
- 06: <http://www.nodes.lk>
- 07: <http://uoj.nodes.lk>
- 08: <http://www.ematnumozihal.lk>
- 09: <http://www.lakapps.lk>
- 10: <http://www.nic.lk>
- 11: <http://www.idns.lk>
- 12: <http://www.nenasa.lk>
- 13: <http://www.Vidatha.lk>
- 14: <http://www.gic.gov.lk>
- 15: <http://www.gov.lk>
- 16: <http://www.documents.gov.lk>
- 17: <http://www.happyliife.lk>
- 18: http://www.jfn.ac.lk/faculties/science/depts/compsc/comsci_glossary/index.html
- 19: <http://www.lakapps.lk/moodle>
- 20: <http://www.icta.lk/index.php/en/e-governement-policy>
- 21: <http://nenasala.lk/pdfdoc/ICTA%20TAMIL%20keyboard%20-%20presentation.pdf>
- 22: <http://www.icta.lk/index.php/en/programmes/pli-development/104-local-languages-initiative-/651-sls-1326-2008-tamil-ict-standard>

E-Governance Initiatives in Tamil Nadu

E.Iniya Nehru, Technical Director,
National Informatics Centre Tamil Nadu State Centre, Chennai
(E-mail: nehru@nic.in)

E-governance Initiatives have resulted in improving core infrastructure and providing better services to citizens. IT has come a long way in changing the lives of the common man at the grassroot level. National Informatics Centre has been providing informatics support to Central Ministries, State Government and District Administration. It has gained deep understanding of governance issues which have paved way for successful designing and implementation of many e-Governance projects like Land Records, Registration, AGMARKNET, Examination Results, e-Post, Passport etc. The impact of government initiatives has been observed in all sectors i.e. administration, agriculture, rural development, judiciary, health, education, telecommunication and transport.

In Tamil Nadu, the major E-Governance projects implemented are explained in detail.

TAMIL NILAM:

Tamil Nadu Infosystem on Land Administration and Management (Tamil NILAM) is a major e-Governance initiative taken by the **Revenue Department of Govt. of Tamil Nadu**. This software system is developed to computerise the Land Records System in Tamil Nadu. *Tamil NILAM* is currently implemented in all the rural taluks in the State. It handles all the transactions relating to Land Records in the State.

???????????? ???? ???? - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited Getting Started Latest Headlines Customize Links Free Hotmail Windows Marketpl... Windows Media Windows
Reliance N... eServices ... NICEHAIL(... INFITT(Int... National I... ICTBook.p... Home: Nat... W E-Governa... Problem I... வட... NIC Tamil ... வட்டா...
எங்கிருந்தும் எந்நேரத்திலும் இணையவழி சேவைகள்
தமிழ்நாடு அரசு
English Version
தமிழ் நாடு அரசின் எந்நேரத்திலும் எங்கிருந்தும் இணையவழி சேவைகளைப் பெற உங்களை அன்புடன் வரவேற்கிறோம். தமிழ் நாட்டிலுள்ள விவசாய நிலங்களின் நில உரிமை (பட்டா / சிட்டா) விவரங்கள் மற்றும் அபதிவேட்டின் படி நில விவரங்களை இங்கு காணலாம்.
இல உரிமை நகல் பார்வையிட...
• இல உரிமை(பட்டா / சிட்டா) விவரங்களை பார்வையிட...
• அபதிவேடு விவரங்களை பார்வையிட...
இணைய வழி இல உரிமை வழங்கியதை சரிபார்க்க
• இல உரிமை(பட்டா / சிட்டா) விவரங்களை சரிபார்க்க
• அபதிவேடு விவரங்களை சரிபார்க்க
• அரசு புறம்போக்கு இல விவரம் பார்வையிட...
Departmental Users Login தமிழ் எழுத்துக்கு இறக்கம் செய்ய
Feedback : eservices@tn.nic.in Help for Installing Tamil Font வகுவாய்த் துறை
Related Links தகவல் தொழில் நுட்பத் துறை
Number of Visitors Since 15th May 2008 : 0514908 தேசிய தகவலியல் மையம்
Done
start jan2010 nehru-ti2010.d... october2006.pdf ICTBook.pdf NICemail - Win... EN 00:48

Objectives of Tamil NILAM:

- Delivery of all possible Citizen-centric e-services
- The entire contents are in TAMIL
- Issue of Chitta Extract (Record of Right) / A Register Extract / Adangal Extract to citizens
- Creation of Master database storing plot wise and owner wise details of land, crop, revenue, etc
- Generation of periodic reports through the computerized system.
- Improved and efficient service
- Easy maintenance and updates of Land Records
- Transparent administration
- Availability of information to public through Touch Screen Kiosk
- Exchange of data to other departments such as Sub Registrar Office, Agriculture Department etc.,

VAHAN & SARATHI:

Vahan and Sarathi are application Software developed for State Transport Authority, Tamil Nadu.



Vahan is for processing all transactions related to Vehicles and Sarathi is for processing Driving Licence and related activities. Vahan can be used to issue of Registration Certificate and Permits. Sarathi can be used to issue a Learners Licence, Permanent Driving Licence, Conductor Licence and also Driving School Licence to the applicant. The system was implemented on pilot basis in in RTO Chennai (North). The system was then approved for implementation in all other RTOs in Tamil Nadu. The systems have now been implemented in 71 offices.

Registration Department (STAR):

Property Registration STAR (Simplified and Transparent Administration of Registration) is implemented in 450 Sub-Registrar Offices. Guideline Values of more than 2 crore Survey Subdivisions are hosted on Internet for public access. The system has enabled citizens to apply for Encumbrance Certificate online. The registration details are maintained in Tamil. The Sub-Registrar Offices and Taluk Offices in 40 taluks are connected using LAN for sharing the data of each other.

EMPLOYMENT ONLINE:

The Professional Employment Exchange Office (PEEO) caters to the employment needs of professionals registered in Tamil Nadu for the entire state. The PEEO office comes under the Directorate of Employment & Training, registers the candidates seeking employment opportunities and recommends to various departments / offices requesting suitable professionals. The PEEO has decided to open their databases for access to the private sector employers, to create more avenues for the registered candidates. The mushrooming growth of private sector employment in this information era necessitates to make online web access of the entire database of registered candidates. In this scenario, the department has created a online web portal for the welfare of candidates and thrown open the entire online database for private sector employment opportunities. The website is operational from June 2003.

The objectives of this web site are as follows:-

- To develop a Data Bank of highly qualified, marketable candidates from the Live Register of the Employment Exchanges in Tamil Nadu with the accent on persons with professional, executive and engineering diploma and degrees.
- To allow the private sector employers easy access to this database to fill up vacancies arising in their establishments and to offer facilities for screening and short-listing of prospective employees
- To provide online information on application dead-lines, track careers and future trends in employment.

Intra Management Information System Portal for Tamil Nadu Pollution Control Board (TNPCB):

This Intranet based Monitoring System has been implemented to create database on all types of Industries (Profiles), the Consents Details and Hazardous Waste Authorisation Details of various types of Industries in Tamil Nadu. This web application is used as an efficient Monitoring System by TNPCB Corporate Office, Chennai. Different types of MIS Reports are being generated using this IntraMIS both at State and District levels. Each TNPCB District Office updates AIR and WATER Pollution related data for all the Industries (taken from the "Application for Consent" submitted by these Industries under Central Act 14 of 1981) functioning in its Jurisdiction. Each industry is identified by a unique File Number assigned by the respective TNPCB District Office. Industry Profile Data, Consents related data and Hazardous Waste related data are being captured for each industry using this web application.

E - Karuvooram : Automated Treasury Bill Passing System:

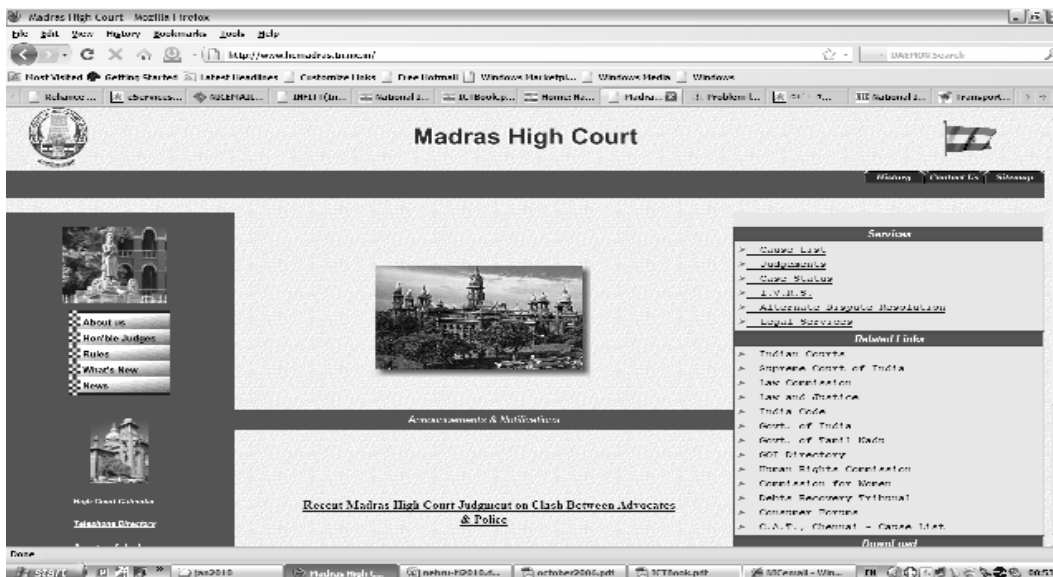
Automated Treasury Bill Passing System (e-Karuvooram) is aimed at automating the existing manual billing system of Treasury Department. The Workflow based Systems developed for the Treasury Department, enables to capturing the data starting from bill submission stage at the counters.

Treasury is the “Bank of Government”, functioning with the objective of maintaining all transactions to the government and sending reports to the Accountant General. Any transaction related to government is performed in the form of a bill. Initially a bill is submitted to the counter of a Sub Treasury / District Treasury / Pay and Accounts Office through a messenger. The bill goes through a phase of approval. This phase is called auditing the bill. Auditing clerk can reject the bill or can approve the bill. If the bill is approved, it is sent to the cheque release counter. The clerks at cheque release counter handover the cheque to the messenger if Treasury type is banking. The application software developed by NIC provides online environment with systems available on all working tables starting from the bill submission counter. The officials can process the bills online and take action for passing / rejecting the claim.

E - Governance at Madras High Court:

Systems in use at Madras High Court deals with

- Posting of Case Status of Madras High Court on Internet
- Display systems are installed at principal bench and Madurai bench to know the status of cases being heard at court halls. 35 court hall display and 6 composite display are installed at Principal bench.
- Certified copies of Final orders, Bails/Anticipatory Bails and interim application orders are entered and issued through the system. The computerized system has reduced the time delay in issuance of copies to the litigants. The Interim application orders are issued through the system since 26th Feb 2001.
- Systems for posting Case Status have been developed. The case details have been hosted on Internet.
- Interactive Voice Response System is installed at Madras High Court to know the Case Status through telephone.
- Touch Screen Kiosk is installed at Madras High Court for public dissemination.
- Daily Cause Lists are being prepared using the system. Systems are installed at Filing Counter, Posting sections for preparation of cause list.



- Daily Cause Lists are being hosted on Internet. Number of hits per day is more than 7000 on working days. As per the statistics available, Madras HC is receiving maximum number of hits.
- Reported Judgments are being hosted on Internet. More than 1500 users visit the Madras HC Judgment site every day.
- Statistical reports relating to disposed cases statements are prepared regularly.
- Information Centre functions at Madras High Court for the benefit of Litigant public to know the status of cases filed at Madras HC. Around 700 case enquiries are received at the enquiry counter every day.

e-Attendance Monitoring System:

This Application is being used by the “e-Governance Cell” of the DoTE, to monitor the Attendance Details of about **2.71 Lakhs Diploma Students** studying in about **340 Polytechnic Institutions** all over the State. Monthly Attendance details of every student are to be entered by the respective Institution in the first week of the following month. Duly signed hard copies of the Branch-wise Attendance Registers are to be generated from this application and to be sent to DoTE for filing purposes.

Any time, a student can login to view his/her Attendance Details for the current Academic Year. At the end of the semester, Hall Tickets will be generated only for the Eligible Students taken from this e-Attendance Database. Thus this Intranet application helps the DoTE to streamline and to bring transparency to the Examination Processing System in the DoTE.

Single Window Counseling System:

This system has been implemented for admissions to more than 390 Teacher Training Institutes in the State.

Common Integrated Police Administration:

The CIPA software is designed and developed to maintain the details pertaining to all the activities of the Police Stations relating to Crime and Criminals. The system provides required information to the higher levels periodically and as and when required. The system also generates various statutory reports for the smooth functioning of the police station. The ultimate goal of the computerization would be an integrated networked system with state of the art hardware and software in place for police access and use the Information in their day to day work and to take decisions. The software is a total work-flow system having the following three major modules along with reports and queries viz., Registration, Investigation and Prosecution.

Agricultural Marketing Information System Network (AGMARKNET):

The AGMARKNET project is sponsored by Directorate of Marketing and Inspection (DMI), Ministry of Agriculture, and implemented by National Informatics Centre. It aims to link all Regulated Markets, State Agricultural marketing Boards / Directorates and DMI regional offices located throughout the country, for effective information exchange on market prices.

The AGMARKNET website (www.agmarknet.nic.in) is a G2C e-governance portal and available in Tamil for the markets in Tamil Nadu, caters to the needs of various stakeholders such as farmers, industry, policy makers and academic institutions by providing agricultural marketing related

information from a single window. It facilitates dissemination, over web, of the daily arrivals and prices of commodities in the agricultural produce markets spread across the country.

PATRAM : Postal Accounts Transaction Maintenance Software :

The PATRAM (Postal Accounts Transaction Maintenance) Software has been designed to maintain all accounting details currently being maintained in various registers. It emulates the auditing functions carried out by the Cash Certificates Section. The PATRAM software consists of 18 modules. This system was developed taking Tamil Nadu Postal Circle as the pilot site and replicated at other sites in the country subsequently.

System for Civil Supplies & Consumer Protection Department :

Family Card Maintenance System

A workflow based Family Card Maintenance System has been developed for the Department of Civil Supplies & Consumer Protection Department, to assist in maintenance of consumer database. Citizens interact with the AC/TSO offices for issue of new family cards, various types of alterations to existing cards, issue of surrender certificates, etc. The software provides a work flow application to manage these services at the AC/TSO offices. All the TSO and Assistant Commissioners' Offices, are provided with internet connectivity. The card database in the central server at the office of the Commissioner of Civil Supplies, is kept updated through online interaction with all the AC/TSO offices. This provides accurate data for computing card-wise entitlement, which is used for preparation of Shop-wise, Taluk-wise, and District-wise monthly allotment statements.

SIM Card Management System(BSNL):

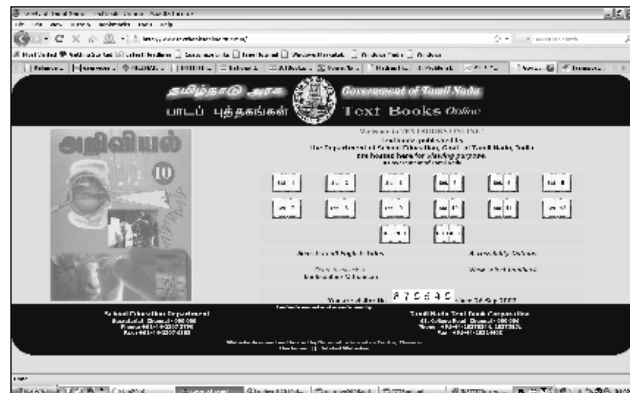
The SIM Card Management System is an intranet application designed - to handle sale transactions at Customer Support Centres; to carry out all activities related to SIM card/ GSM number distribution/ allotment by the headquarters; and to maintain accounts; through a centralised database. The SIM Card Management system uses intranet technology to give a cost-effective solution that delivers powerful and robust results. By connecting Customer Support Centres, dealers, and distributors with the Mobile Services headquarters, through a centralised server, information is managed effectively and efficiently.

Chennai Corporation:

Ten Zonal Offices of **Corporation of Chennai** make use of the intranet based systems developed for Property Tax collection, Birth/ Death Certificates extraction, Company Tax collection, etc by accessing the central server located at Head Office.

Text Books Online:

The entire text books of school children under School Education Department, Tamil Nadu Government is available online.



Geographical Information System (GIS):

Web GIS tools under Open Source environment have been used for developing the TN Maps site <http://tnmaps.tn.nic.in>. The tools have also been used for generating dynamic maps using Census 2001 data at <http://www.census.tn.nic.in>.

E-filing of Returns by Dealers for Department of Commercial Taxes:

All the dealers in Tamil Nadu are filing their VAT monthly returns online through the website.

Digital Certificates:

PKI enabled application systems were developed for Madras Export Processing Special Economic Zone (MEPZ SEZ) to facilitate exporters registered with the Zone to file their Applications and Quarterly / Annual Returns through the Internet.

Web Services :

The website of Government of Tamil Nadu is being maintained by NIC and has large number of citizen oriented particulars such as Policy Notes of all the departments, Citizen Charter, RTI documents, Government Orders of Public Interest, Public Utility Forms, Tender Notices, Press Releases, Contact details, Statistical Reports, Acts and Rules documents, etc. The websites of all the districts and those of many departments have been developed and hosted by NIC.

The GPF particulars of more than 5 lakh employees of the State are hosted on the web through the website of Accountant General of Tamil Nadu. Pension processing status is also made available through the website.

Online Registration System is used for the examinations conducted by the Tamil Nadu Public Service Commission. More than twenty thousand applications have been received online in the last two years for various examinations.

Electoral data :

All the 234 Assembly Constituencies are hosted on the web with interface to search the data in Tamil.

Examination Results :

The examination results of the Teachers Recruitment Board are hosted on the web. Tenth and Twelfth Standard examination results were accessed by more than 20 lakh users are hosted every year from 1997 (one of the first States to have exam results online).

E-Governance is a continuous journey and Tamil Nadu has been an active State in this mission mode journey.

E- Governance in Tamil for Tamil Virtual University

G. Amirtharaj, Software Engineer

Dr. A James, Consultant

Dr. P R Nakkeeran Director

Tamil Virtual University, Taramani,
Chennai - 600113, Tamil Nadu, India.

Email: tamilvu@vsnl.com

Abstract:

Information and Communication Technology (ICT) has today become an integral part of governance, especially in India. ICT is viewed as a tool that will help deliver services in both public and private sector faster and transparent to end users.

E-governance, as a concept, involves leveraging ICT to streamline the administrative process. It involves computerization of records, facilitating efficient transactions between various departments using web portals and other electronic data transfer mechanisms to bring administration more effectively, to achieve the goals faster and ease.

Around the world various private, public and Government sectors are implementing the e-governance to make their process transparent, cost effective, providing any where any time access of services and to reduce the processing time of their work flows.

In order to make the process transparent, cost effective and to reduce the processing time Tamil Virtual University Management has decided to implement the **E-Governance in Tamil Language** for various activities like Employee Management, Time Management, Leave Management, Vendor Management, Work flow management, Course management, stock management and etc...

Introduction:

This paper speaks about the implementation of E-Governance web based application in Tamil language for TVU with the Tamil Unicode support. Tamil Virtual University (www.tamilvu.org) is a Tamil Nadu Government Organization aims at providing Internet based resources and opportunities for the Tamil communities living in different parts of the globe as well as others interested in learning Tamil and acquiring knowledge of the history, art, literature and culture of the Tamils. The functions of TVU include Internet based Educational Programs, Digital Library and Development of Tamil computing.

As part of e-governance initiative, the TVU management embarked on a paper-less office drive for introducing transparency and accountability in its internal and external transactions with vendors. As an initiative TVU has started developing a web based platform independent e-governance tool using Java, Struts Framework and MYSQL database with Tamil Unicode support.

Towards Implementation of e-governance applications for TVU:

This e-governance application is a web based application developed using Java Technologies (JSP and Servlets) and Struts Framework, MYSQL database as backend, Apache web server and Tomcat application server running in Linux OS with Tamil Unicode Support. The backend database is developed using MYSQL. It is configured to store Tamil Unicode Characters. This application allows to key in Tamil & English characters by toggle between English and Tamil language. The data transactions between client and server are secured by Secured Socket Layer (SSL) enabled Apache web server.

Application Architecture:

The below diagram (Fig-1) describes the architecture of the e-governance application for TVU.

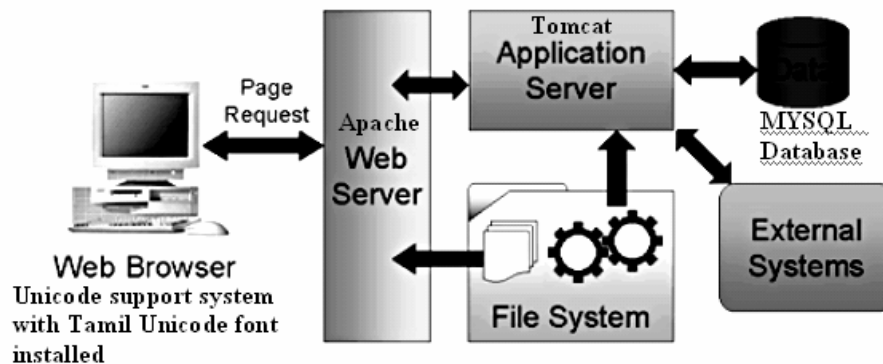


Fig-1 – Application Architecture

Web Browser – Unicode (UTF-8) supported system with Tamil Unicode font installed to view the application pages by the end users. **Apache web server** – to serve the html, images and media content. **Tomcat Application server** – java container to run the JSP/Servlet. All the requests/response should be receive/sent in UTF-8 encoding format to support Tamil Unicode encoding scheme. **MYSQL database** – UTF-8 supporting database tables to store Tamil Unicode characters. **File System** – Configuration Management System to store electronic forms of documents/files. **External System** – Online Payment System (Currently ICICI payment gateway).

Now we will look into the details of each application:

Gateway Portal:

This web portal is the gateway to access all the applications in a single window. Each entity i.e. employees, study centre admin, vendors and students will be created with unique user name and password through user management module. An authorized admin will give access to create/update the user details.

The user can login into this portal and access the different application modules based on their role. This portal can be accessed from anywhere at anytime through web browser with internet connection enabled and also through intranet.

தமிழ் இணையப் பல்கலைக்கழகம்

பயனர் உட்புகுதல் User Login

பயனர் முகவரி User ID:

கடவுச்சொல் Password:

Fig - Login screen

Employee Management:

This application is to manage the TVU employee details and Personal Registry (PR). Admin person can add new employee and update the user detail through the CRUD (Create Read Update Delete) process. A user (TVU staff) login into the e-governance portal can view their details like their personal detail, skill detail, years of experience, designation and his manager hierarchy. The employee will be allowed to update his personal details only.

தமிழ் இணையப் பல்கலைக்கழகம்

முகப்பு கடவுச்சொல்

வணக்கம் amirtharaj!

விவரம்:

பெயர்:	திரு. அமிர்தராஜ் கோ.
பணி. எண்:	133
பதவிப் பெயர் :	மென்பொருள் பொறியாளர்
பி.நா:	11-March-1976

நே. மேற்பார்வையாளர்:
முனைவர் ஜேம்ஸ்.அ.

முதன்மை. மேற்பார்வையாளர்: இயக்குநர்

உதவியாளர்:

படிவங்கள்

தன் பதிவேடு(PR)
விடுப்பு
Extra Working
நகல் அச்சு

மேலாண்மை

விடுப்பு அறிக்கை
தன் பதிவேடு(PR) அறிக்கை
பணி நிலை அறிக்கை

ஒப்புதலுக்காக

உங்களுக்கு உதவியாளர்கள் இல்லை!

Fig- Employee home page

In TVU, personal registry was maintained by entering their daily activities manually in a physical notebook, routed to the reporting authority through attendant and verified. E-governance facilitates online filing daily status, after implementing this module the employee will enter their work details in the online form and submit will intimate their immediate managers through mail to give their approval. The managers will login to the portal and view the pending PR's for approval. The manager can give their approval through a single mouse click after verifying the details. Here the employee can generate their work report of his/her own or his/her sub-ordinates. The reports can be generated monthly wise or a whole.

Time Management:

As part of the e-governance, TVU has implemented the thumb impression attendance machine (Bio-metric system) from a third party. Until recently the TVU staff has to put their signature on the physical attendance book daily but it will not have the details of in-time or out-time.

தமிழ் இணையப் பல்கலைக்கழகம்

முனையு ல் லுமொலகை ன்றது வெளிடு

தன் பதிவேடு(PR) அறிக்கை

ஒரு பஸியாளரின் தன் பதிவேடு யாத அறிக்கையை தோற்றுவிக்க, அந்த பஸியாளர் பெயர், தன் பதிவேடு யாதும் யற்றும் வருடத்தை தோவு செய்து, தோற்றுவிக்க பொத்தானை அழுதலும்.

பஸி. பெயர் : திரு அனந்தராஜ் லீலா

யாதம் : லா

வருடம் : 2010

தோற்றுவிக்க

பின்

Fig - PR Report

For payroll calculation, a clerical person has to verify manually for each month and based on that monthly salary will be calculated. After implementing the Bio-metric thumb impression attendance machine now the employee has to give his/her thumb impression while entering or leaving the main entrance every time. The data will be collected in SQL server database with all the in-time and out-time details. A standalone reporting tool has been provided to generate/view the reports.

Report

1 of 2 100% Find | Next

தமிழ் இணையப் பல்கலைக்கழகம்

Staff Summary From Apr 26 2010 To Apr 26 2010 Printed: Apr 27 2010 11:09

Location: chennai Department: STAFF MEMBERS OF TVU

Staff Name: முனைவர். நகீரன் ப.அர். Staff Code : 101

Sl.No	Date	Shift Name	Duration	Late By	Early Going	OT	Status
1	Apr 26 2010	General	8:09	00:00	00:00	00:00	Present
			8:09	00:00	00:00	00:00	

Staff Name: முனைவர். நைச்சாமி முகமது இக்பால். சே.பி. Staff Code : 110

Sl.No	Date	Shift Name	Duration	Late By	Early Going	OT	Status
1	Apr 26 2010	General	00:00	00:00	00:00	00:00	Absent
			00:00	00:00	00:00	00:00	

Staff Name: முனைவர். சித்தலிங் கையா, தி.பி. Staff Code : 111

Sl.No	Date	Shift Name	Duration	Late By	Early Going	OT	Status
1	Apr 26 2010	General	8:01	00:00	00:00	00:00	Present
			8:01	00:00	00:00	00:00	

Fig - Attendance report

Every two weeks reports will be generated and sent to all the employees through mail. In the near future, the SQL server database will be integrated with our e-governance portal so that each employee can generate their reports of their own or for their sub-ordinates and can view through the portal itself. And also the accounts department can get the consolidated month wise report of all the employees.

Leave Management:

Through the e-governance portal the TVU employees can login and apply their leave through online, by selecting from and end date (including a half day), reason for the leave, type of the leave (casual, medical and earn leave) and submit the leave form. The managers will login to the portal and view the pending leave forms for approval. The manager can give their approval through a single mouse click. The employee can apply/approve the leave from anywhere at anytime. This application will be integrated with the Bio-metric attendance machine database so that if an employee forgot to apply leave forms will automatically assigned leave for that particular day. Through this application and authorized admin person can manage the government and other holidays so that this will be reflected in the "MYCLANDER" option for each employee to view the holiday details. Monthly attendance and leave reports for the employees can be generated and other various individual employee reports in Tamil.

Vendor Management:

This application is envisage to manage the TVU vendors, project details assigned to the various vendors and quotation/tender details. Through this application a TVU vendor can login to this portal through his provided username and password and can view project details assigned to different vendors, its completion details etc... Here an authorized TVU staff will create online work order form and fill up with the project details like project title, vendor name, starting and end date of a project and submit will go for the approval to the concern approval authority, once it has been approved a final output will be taken with the authorized signature and provided to the vendor. The approval authority will assign a particular project to a TVU employee to follow-up the project. Now that employee and his sub-ordinate will get the full details of the project in the e-governance portal though his login. The employee can track the project and follow up the vendor until the project completes. All the communication, demos, project status, vendor visiting detail, CD's submission etc... will be stored corresponding to this project and the higher authorities can the view the details at anytime. A detail report of project details and status of the project can be generated based on different criteria like project wise, vendor wise etc...

Work flow Management:

Through this application the TVU staff members can submit the online Xerox/Print. Only after approval from authorities regarding details like number of copies and etc...actual printing takes place. The admin will take the Print/Xerox and update the form with actual number of copies taken and submit will go the supervisor for approval. The admin can generate the Xerox/print taken reports based on different criteria like for a particular month, by an employee, on particular device etc...

Course Management:

This application is to manage TVU Study centre across the globe, student, providing permission to access the online examination for a students, generating question papers for online and offline examinations, conducting or monitoring of online examination, mark sheet and certificate generation for students and examination result publishing. The study admin can login to the portal and allowed to register a new student for a course, to pay students course fee, register for a examination, providing permission for offline and online examination for their students, password reset etc...

Fig – Xerox/Print Request Form

TVU student registered for a course will be provided with unique username and password and the student can access the portal to view the examination schedule, online examination samples, question pattern samples for both online and offline examinations, exam results and mark sheet.

TVU examination controller or the authorized TVU employee can register a new student for courses, register a student for examination, password reset, publishing examination results, generating examination report and sent to the study centre.

Physical Stock Management:

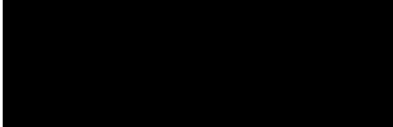
Now in TVU assets like furniture, computers hardware and software, Physical library books and course materials, Course CD's, Electrical accessories, paper and other office accessories are maintained in the physical register. The existing stock management is very cumbersome and difficult to integrate purchasing of available stock with consumables.

Through this e-governance portal an interface will be provided to the authorized TVU staff to add new asset through the online form in different category like furniture, electrical, computer etc..., of type consumable and non consumable and with other details like date of purchase, make, vendor and support/service contact details, warranty and location details. After submitting the online form an asset ID will be generated. Then the higher authority will verify and give his approval. The new items will be tagged with the generated asset ID. If a non consumable is abandoned or consumable has been used, through this application interface the details will be updated for that particular asset and

submit. And this application will allow authorized TVU staff to generate the report based on various criteria.

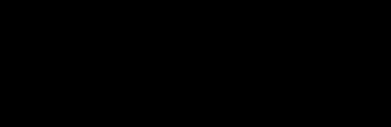
Conclusion:

This paper has discussed about the implementation of e-governance web based portal in Tamil language for TVU with the Tamil Unicode support. By implementing the e-governance application in localized Tamil language using Tamil Unicode encoding scheme, it will remove the language barriers in implementing e-governance application in all the Government departments to facilitate efficient transactions between various departments and public using web portals and other electronic data transfer mechanisms through single window to bring administration more effectively, paper-less office drive and to provide effective services to the public.



13

கணினி வழி கல்வி



E-Resources are the best Information Service to Teach, Learn and Research through World Wide Web

V. Thangavel

Research Scholar

Dept of Library and Information
Services
SCSVMV University
Kancheepuram - 631 561

D. Mohanraj

Professor

Dept of Management Studies
Paventhar Bharathidasan
College of Engg and Tech
Trichy.

Dr. Ramesh

Professor & Head

Dept of Chemistry
Kalsar College of Engg.
Sriperumpudhur
Chennai- 602 105.

Abstract

The paper highlights the usage trends of access to e-resources in Indian Universities, Colleges and Research centre. The preliminary findings proceedings abstract of various conference of last ten years revealed that there is an upward trend. The paper briefly describes open access e-resources used by the various scholars through World Wide Web in India: A citation Analysis

Key Words: E-Resources, Consortiums, Information Sharing, E-Books, E-Journals, E-Magazine, E-Thesis, E-paper, E-Library, E-Publishing, Digital or Virtual Library, E-Lifestyle, E-Government, E-Directories.

Introduction:

World wide web has created a sea change in providing information and as well as information transfer, Information scientists have undergone a perceptible change and have become IT centered, which is evinced from the publications pattern of articles, citations in journals, seminars, conferences ect., The world wide web provides access to materials that were previously not available to the researchers, students and faculty now able to view, listen to, and read materials that just few years ago were difficult, if not impossible, to access. Images, texts, historical documents, video clips, sound files etc, are now available over the web to anyone that has internet access. Almost any one can put almost anything online, and because that information is not filtered or mediated in any way. Academic, Research and development organizations are supported by well equipped library and information services unit to augment the objectives of the parent organizations. Proliferation of knowledge growth has resulted in information explosion, which has made the users so difficult to find the relevant information of research and development.

The modern libraries are able to provide information services to their patrons more efficiently with applications of information communication technology (ICT) in their environment, which satisfy the currency of information. In the recent years, the publishers and the professional societies are able to add value to their publications by means of providing the full text contents in electronic form. The networked

environment of the organizations and the availability of internet connectivity give the users and easy access to the full text of the document.

Purpose:

The purpose of present paper is to determine the awareness of electronic resources among library users and promote the usage of electronic resources by the students and members of faculty in engineering colleges in Tamil Nadu. Questionnaire was distributed to 210 randomly selected faculty and students in the various research institutions in Tamil Nadu . The researcher collected many articles i.e. print and on-line. Nearly 54 articles have been received. Through this articles we identified the use of e-resources for various purpose i.e. teach, learn and research. Majority o the user using for research purpose. The 40% of user using for teaching purpose. The considerable percentage of users is using the e-resources for learning purpose. Through this articles reveals that the researcher and faculties are using e-resources through www, so the user's categories also increase day by day. Because the Government also to allot more fund innovative consortium for the user categories. In future e-resources are a pillar for servicing various information.

Retrieval Systems:

The study is an attempt to determine the level of using various types of resources by the research scholar in various University, Scholars felt that about various issue surrounding the electronic resources and whether changing attitude depend upon subject. Further reports have been presented about the factors supporting the growth and academic work with the help of web. The world of academics enjoying the fruits of electronic resources by the way of resources retrieval is highly thankful to the person like Tim Bern Lee for their tremendous effort. The 50% web resources as noted by the respondents gives the positive signs of the growth of IT world in the academic venture. Though the users of the web are having basic knowledge to handle the electronic resources, the institution involved in academic work should provide technical training to their students. The free internet access available in most of the universities may be retained and may be extended to other universities also. The mushrooming of new website development may lead to unreliable resources to the students at large and the research scholars in particular, hence, a board may be constituted to check the validity of resources before it is uploaded in to the web.

E-Resources:

E-Publishing: Electronic publishing (e-publishing) is 16 years old and the first e-book was published in Germany in 1985. The products of e-publishing are mostly secondary sources, reference materials and the primary periodicals. The problems associated with e-publications are integration with traditional forms, cost of acquisition, collection development, non-availability of selection tools, etc. These challenges should be taken as opportunities and prepared to find out the standards. When the **father of the printing press, John Gutenberg**, looked towards the future, he could never have dreamed that one small invention could revolutionize the world. The gift of publishing benefited not only the authors who could now create new words, but also the readers who would be exposed to new ideas, and new concepts. The creation of the printing press sparked a new revolution-a literary revolution. Now, with the widespread availability of computers and the internet, technology, enables virtually anyone to self-publish their literary works. E-Publishing is primed to revolutionize the world, by giving every individual the ability

to publish their ideas, stories, books, without the prohibitive costs associated with conventional publishing. E-publishing is a process for production of typeset-quality documents containing text, graphics, pictures, tables, equations etc. In general, it is used to mean any information source published in electronic

E-Books:

Libraries form a vital part of the world's systems of education and information storage and retrieval. They make available – through books, films, recordings, and other media-knowledge that has been accumulated through the ages. People in all walks of life-including students, teachers, business executives, government officials, scholars and scientists – use library resources in their work. Large numbers of people also turn to libraries to satisfy a desire for knowledge or to obtain material for some kind of leisure-time activity. In addition, many people enjoy book discussions, film shows, lectures, and other activities that are provided by their local library. There are a considerable number of concepts being explored in future libraries research and development. The focus is in distributed and local collections of information objects – in the hybrid library including analogue as well as digital objects and on ways of identifying objects of interest to a user and arranging for the user to access them. The library of the future will be less a place where information is kept than a portal through which students and faculty will access the vast information resources of the world. The library of the future will be about access and knowledge management, not about ownership.

E-Journals:

Advancement in science and technology has changed the knowledge communication in the form of e-media. Order of the day is e-publishing and tradition print sources are replaced by e-sources. An attempt has been made to study the utilization and satisfaction among the users of various institutions and research centers towards e-journals using survey. The modern society is dynamic and complex. The duty of the research scholar towards social change, scientific development and social uplift is undisputable. E-Journals will definitely made its own impact on the users in terms of scholarly communication of research.

The change in information technology and digital library has made us to move from print to electronic media in terms of acquiring information resources and to provide services. Libraries are undergoing rapid changes due to the developments in information communication technology paper based resources are giving way to electronic resources. The CD ROM and E-journals are providing information to the user from various research communities. An attempt has been made in this study to identify the usage of CD ROM s and E-resources among the various research community in Tamil Nadu.

E-Government:

The term “e-government” is used in this paper to denote the concept of using information communication Technology (ICT) as a means to organize and manage the administrative processes of the government, especially the interactive processes between the government and the public. Though ICT has been available widely for more than four decades and many governments around the world have indeed used ICT in certain aspect of government, the concept of e-government is relatively new in the sense mentioned above. Only a handful of governments have progressed to a high degree in harnessing the immense power of ICT in re-organizing their government infrastructure and in serving their citizenry,

and have done so in an efficient and effective manner. E-government is not mere “technologising” of government. It is not just a matter of automating some manual processes nor is it a simple introduction of technology where none existed. E-government requires a fundamental re-thinking of governance itself and, as some have suggested, re-inventing of government. If bureaucracy is the invention of the 19th century, we might say e-government is the invention of the 21st century. E-government re-examines the organizing principles of bureaucracy and governance, re-defines the objectives and deliverables of government and re-deploys the resources available. In this process of re-invention, the basic intent is both refinement of the old and introduction of the new. E-government is NOT throwing the baby with the bathwater.

Consortiums for E-Resources:

Consortium a strategic alliance of libraries with a common interest, not under the same institutional control, but usually restricted to a geographical area, number of libraries, types of materials, or subject of interest, which is established to develop and implement resource sharing among member. This paper discussed about the benefits and limitations of e-resources through consortium. Makes a comparative study of two major library consortium of India i.e. INDEST-AICTE & UGC- INFONET in terms of objectives, governance, members, access pattern, resources etc. Analysis the study and makes suitable suggestions for its improvement. Concludes with the remarks that these consortiums will bring remarkable change in the library scenario and also the educational system of India. In modern library environment, information technology is playing a pivotal role. Today’s technology, interactivity, digital media, expanding network and communications capabilities compel the libraries to move from organizational self-sufficiency to a collaborative mode of acquisition of library materials. Library consortia are one of the emerging tool kit. The consortium should take a lead role in the development of a national strategy for information provision for research in higher education. Librarians should seriously rethink and reinitiate consortium movement like developed countries for maximum utilization of resources at a reduced cost, time and space. It is an encouraging sign that both INDEST-AICET and UGC INFONET consortia are functioning well. These will bring remarkable change in the library scenario as well as the educational system of India.

The UGC - INFONET programme will be a boon to higher education systems in several ways. It is a vehicle for distance learning to facilitate the quality education all over the country. It is a major source for research scholars for tapping the most up-to-date information. It is a medium for collaboration among research guide and research scholars. One of the major developments in libraries and information systems in the past 15 years is the advent and spread of electronic information resources, services and networks mainly as a result of developments in information and communication technologies. The change is basically of physical form where information content is increasingly being captured, processed, stored and disseminated in electronic form. The unique features of the information needs of users in electronic environments relate to the physical form in which information content is made available in electronic information environments. Users normally desire the content to be made available within the constraints of their skills and technological capabilities so that it is possible to access and use the required information content to resolve the felt gap in knowledge.

The E-Journal programme is corner stone of the UGC-Infonet effort which aims at addressing the teaching, learning and research collectively and governance requirements of the universities and research

centers. It would facilitate free access to scholarly journals and databases in all areas of learning to the research and academic community.

This article discusses the need for sharing of electronic resources among Libraries and Information Centers in developing countries. Highlights the importance of 'Library consortia' in this digital era, stating its salient features, functions and responsibilities. Examines in detail the emergence of various resource sharing models in the developing countries with special reference to India.

Emergence of Library Networks in India:

In India library networks actually started due to the initiatives by NISSAT in forming CALIBNET in 1986, DELNET in 1988 and another networks subsequently. The UGC (University Grants Commission, India) set up INFLIBNET in 1988, NIC (National Informatics Centre) also set up DELNET, which is the first operational library networks in India. CALIBNET and INFLIBNET have not been able to project their performance as they were planned but they are trying to achieve their goal. Functions, organizations, cooperation's, progress and creation of databases amongst libraries of BONET and CALIBNET as library networks are still dissatisfactory. Institute of Economic Growth (IEG) was founded in 1958 as an autonomous institution recognized by the University of Delhi. The IEG library has its networking with different libraries e.g. Ratan Tata Library, Delhi School of Economics, Delhi University library, NCAER library, IIPA Library, DELNET etc. for helping the readers in getting the desired books and journals which are not available in the IEG library on inter-library loan basis. There are other examples of networking and resource sharing of Astronomy libraries in India. They jointly established a networking for resource sharing amongst the libraries e.g. Indian Institute of Astrophysics (IIA) library, Inter-University Centre for Astronomy and Astrophysics (IUCAA) Library, National Centre for Astronomy and Astrophysics (NCRA) Library, Nizamiah Observatory (NO) Library, Physical Research Laboratory (PRL) Library, Raman Research Institute (RRI) Library, Tata Institute of Fundamental Research (TIFR) Library, Uttar Pradesh State Observatory (UPSO) Library. The main purposes of this networking are for better resource sharing, to reduce costs, for speedy delivery of documents, to keep abreast of new developments etc.

No efforts have been made in India to network public libraries though it is very much essential to cater networked information to the general public where more than 70% of the total people are residing in rural areas. However, much emphasis should be given at the national level in India for the development of documentary information resources because it is considered as one of the vital resources to promote the development of economy, science, technology and culture etc.

Resource Sharing Networking Models:

The fundamental object of information resource sharing is to solve the problems faced by the libraries such as information explosion, several needs of users, increasing cost of subscribing periodicals, sharing library budgets, fluctuation of the exchange rates etc. Now-a-days, to solve the above problems different resource sharing networking models are observed at local, regional, national and international levels. Generally, there are three levels of national resource sharing networks exist: (a) Local: Information is stored in local libraries in the form of Union Catalog for local collection available in local libraries; (b) Regional: Information is stored in regional libraries and services are provided on broad subject area basis; and (c) National: Information is stored in national library in the form of national Union Catalog for

normally national collection available in national library and services are rendered to users as national resources. There are four existing resource sharing networking models, which have been shown in tabular form in Table – 1. VARIOUS MODELS TO RESOURCE SHARING

Model	Aims & Funding	Examples
Centralized collection development and Services at national and Regional level.	Resources: Acquired centrally and stored at single site Funding: Contribution by participating libraries. Grants are also sought from government and private agencies.	National Lending Library. UK.
Centralized Collection Development and Services by Subject	Resources: Subject specific collection of documentary resources. Acquired centrally and stored at a single site city, region, or country may limit the geographic distribution of libraries. Funding: Marketing of services and grants from the government and private agencies.	National Service Library at INSDOC. New Delhi
Centralized Collection Development at Organizational level	Resources: Libraries belonging to a single bigger organization collaborate. The shared collection is acquired centrally at a single site. Funding: Organization backing the library provides funds. The Participating libraries may also contribute towards the central funds.	CSIR, DRDO, DOE, ISRO
Coordinated collection development at Institutional level	Resources: Eliminates duplications. Serves at the level of participating libraries. The geographical area of cooperation could confine to a city, region, or country. Funding: The individual libraries determine their level of support. User libraries pay for the services they avail of.	DELNET, BONET, MALIBENT

UGC –Infonet Electronic Journal Consortium is an innovative project launched by University Grant Commission (UGC) to promote the use of electronic data bases and full text access to journals by the research and academic community in the country. This project will bring about a qualitative change in the academic environment. The research and academic environment. The research and academic community can now have an access to resources at their finger tips.

Consortia Initiatives in India:

The information revolution and proliferation have brought about drastic changes to the function and service in all types of libraries in India during last two decades. Many libraries in India till today are not in affording position to procure all documents and subscribe to core journals in major disciplines or CD-ROM databases, due to their financial crunch. At the national level, the UGC (University Grant Commission, India) setup INFLIBENT in 1988. It has taken imitation about a formidable change in developing adequate infrastructure in libraries, especially university libraries, to be a part in the

networked environment. Since January 2004, University Grants Commission through its one point programme (UGC) / INFONET is providing access to e-subscription to all-important journals for the entire university community. The Ministry of Human Resource Development (MHRD) through its INDEST (Indian National Digital Library in Science and Technology) has launched consortia-based subscription to electronic resources for higher – technical education systems in India.

Besides, there are a few national and regional library consortia developed in the recent years. Council of Scientific & Industrial Research (CSIR) Tata Institute of Fundamental Research (TIFR),, Department of Atomic Energy (DAE), Indian Institute of Technology (IIT), and Indian Institute of Management (IIM) have already formed their sectoral consortia and have been subscribing to electronic sources like Science Direct, MathSciNet, and Blackwell, John Wiley, ABI/INFORM and Business Sources Premier. Also, both Institute of Mathematical Science (IMSc) and TIFR have been subscribing to MathSciNet database under their own consortia consisting group of libraries in their region. And many more are expected to come soon.

Increasing the Discovery and use of e-resources in University Libraries and Research Centres:

There is a large quantity of subscribed e-resources in our libraries and they contain quality information, though expensive. In spite of advantages in terms of access and search capabilities, they are underused. Systematic plan has to be in place for their promotion of use. While a good ICT infrastructure is a prerequisite, it alone will not do. Proactive strategies are required and these need to be adopted imaginatively. Access to e-resources need to be made easier for both on campus and off campus users. As a priority, active users need to be identified and they need to be converted to heavy users of e resources. Secondly, non users be converted to active users Various methods have to be tried in order to grab the attention of the users towards the e-resources. User training will increase the confidence level of the users. Traditional awareness methods include : Personal visits, user training, brochures, posters and displays. Newer technologies from the Web 2.0 such as RSS alert service, Blogs, Wikis and Facebook make the interaction with the library not only interesting but also add more value. Finally, the effectiveness of various promotional strategies need to be measured by monitoring the usage and user feed back. This research sought to determine use of online resources and databases and to assess current user characteristics associated with use of online resources by the Faculty and researcher of various universities

Suggestions:

1. The university library should make necessary arrangement for continuing subscription for the on line journals along with print journals.
2. The University library should conduct orientation, training programme regularly to assist the users of agricultural consortium programme.
3. The authorities of the university should take keen interest for providing better infrastructure facilities for the improvement of Internet speed, e-journals, e-books, e-directories, e-conference proceedings, e-thesis, e-dissertation and e-dictionaries. So that users can feel more comfortable in browsing on line information.

Conclusion:

Remote access of e-resources has been a major boon to academic and research libraries. Online journals are considered the note chord of any library's collection and have become indispensable for research in any field. Many online journals available in the form of databases as well as they directly access through the Internet. The quantity of online journals is growing larger and has become a quite visible entity in serial publication. Today most of the online journals appear as parallel version of its print counterparts and more publishers are making their journals available in electronic format. Many academic institutions are currently building substantial collections of full - text journals and continue to increase access to various online databases. Because these resources come at a great cost, it becomes important to understand database and full-text journal use among university patrons and the characteristics accompanying today's remote and in-house library users. Increased access to computers, the Internet, online databases full text journals necessitates reassessing online use patterns and user characteristics. Nowadays it is impossible for libraries to procure all the documents and subscribe to core journals that are in demand by the users. There are many online journals and databases are available open access. Subscription of online journals and databases through the consortium(s) are much economic for the libraries.

Reference:

1. ISI Reseachsoft.2003a.EndNote-Product Information, 5 August 2002, [Http://www.endnote.com/eninfo.asp](http://www.endnote.com/eninfo.asp)
2. Asksamsystems.2003,cittion Bibliographic and Research Note Software, 15 Feb 2003, <http://www.citationonline.net>.
3. Craztsqyurrek cinoekte siytuibs, 2003 Refas – The Reference Assistant, 2002,
4. Scholars aid inc.2003 is scholars aid right for you?, 27 aug 2002, http://scholarsid.com/right4you_research.html
5. Brown, C.M. 1999, "Information seeking Behaviour of Scientists in the Electronic Information Age: Astronomers, Chemists, Mathematicians, and Physicists: Journals of the American society for Inforamtion Sceince 50 (10): pp 929-943.
6. Green, R (2000) Locating sources in Humanities scholarship: The efficacy of following biblio graphic references, Library quarterly 70 (2) pp 201-229.
7. Bates, M J (2002) The cascade of interactions in the digital interface. Information processing & management, 38 (3), 381-400.
8. MathesA (2005) Preserving Public Domain Books. <http://googleblog.blogspot.com/2005/11/preserving-publicdomain-books.html>
9. Barrett A (2005) The information seeking habits of graduate student researchers in the humanities. The journal of Academic Librarianship, 31(4), pp.324-331.
10. Workloc, Kate (2002) Electronic Journals: User Realities – The Trusty about content usage among the STM community, Learned publishing 15 (3): pp.223-226.
11. <http://www.digitaldivide.net>
12. <http://www.lisnews.org/~jay/journal>
13. <http://allrss.com/rssreaders.html>
14. www.weblog.com/2005/v2nl/a10.thml.

Use of E- Resources by the Research Scholars and Faculty of Management Studies in Management Research Centres, Tamil Nadu

V.Thangavel

Dept of Library and Information
Services, SCSVMV University,
Kancheepuram – 631 561
estv_libn1@yahoo.co.in

Dr. K.Ramakrishna Reddy

Dept of Management Studies,
Paventhar Bharathidasan College
of Engg and Tech
Trichy

Dr. V.Ramesh

Dept of Chemistry, Kalsar
College of Engg. Mannur,
Sriperumpudhur
Chennai- 602 105.

Abstract

The electronic resources have the effect of democratizing the research community. The study is an attempt to determine the level of using various types of resources by the research scholar. Research scholars felt that the electronic resources would be useful to carry out their research and do their further research works, depending on their subject. This article discusses about the use of electronic resources by the research scholars and faculties of management studies through World Wide Web in Chennai, Tamil Nadu.

Introduction:

Computer-based automation was initially incorporated into library operations as a mechanism for handling the routine functions of running a library such as circulation, cataloging, acquisitions, interlibrary loan and serials control. Systems for handling these operations became available able to the larger library community from the early 1970s onward, although there were some earlier pioneers with well developed local systems. Early systems were typically run on large computers into which data were entered, processed behind the scenes, and returned as printed output of some type (overdue notices, invoices, or catalog cards, for example).Pioneers in reference database services, such as DIALG, also operated using large mainframe computers to which patrons connected via terminals and required experienced searchers mastering a complex set of commands to generate effective search results.

The modern society is dynamic and complex. The duty of research scholar should know the availability of sources in their topics. Electronic resources are available in various formats. The different types of electronic resources are electronic books, electronic journals, electronic databases, search engines and subject gateways.

Public Services:

Electronic resources are changing the longstanding relationship which traditionally existed between library professionals and users. Formerly, the use of resources and services required patrons to visit the library. Print indexes in the reference area provided needed citations the card catalog indicated whether that library owned a particular item, and the circulation desk could place a hold for an item checked out to another patron. Early electronic tools such as DIALOG were available only to librarians who

frequently consulted with users to refine their needs before conducting expensive and sometimes complex database searches. The introduction of large scale automated union catalogs like OCLC have helped to revolutionize services such as interlibrary loan and library functions such as acquisitions and cataloging. The introductions of newer electronic resources have shifted services and functions from those which are librarian mediated to those specifically geared towards the end user. Advances such as the availability of online citation and document delivery services remotely accessed from a home or voice, 24 hour remote availability to OPACs providing both an item's current circulation status and the ability to place holds oneself, plus the increasing number of resources available directly to patrons through networks, have combined to eliminate the need for a visit to the reference desk, or the library in many cases, to satisfy information needs.

Objectives:

The following objectives are framed for study.

1. To evaluate the purpose of using electronic resources.
2. To identify the use of electronic resources.
3. To identify the purpose of electronic resources by the research scholars and faculty in various field in Management Research centers, Chennai, Tamail Nadu.
4. To find out the best publicity method to promote the usage of electronic resources.
5. To identify the format of print or electronic resources and their performance.
6. To know the problems faced by Research Scholar in accessing electronic information.

Literature Review:

A good number of related studies on the usage of electronic resources have been conducted mainly in developing countries like India and developed countries. These studies led to know the status of the usage of the electronic resources against investment.

Dugdale studied the library services at the Bolland library, university of the west of England, Bristol in the UK and investigate the ways; in which students might be encouraged to use electronic resources and to develop important life long learning skills through the Reside (research, information, delivery) electronic library.

Ferguson stated that the university of Hong Kong in China had found that the change to a mostly elec5tonic collection has been successful. He stated that 59% patrons preferred reading electronics serials while 30% favored the printed version. Only 14% of the academic staff still favoured printed journals.

Alwarammal R , Sivaraj S and Madasamy R stated that respondents were asked to mark their potential use of electronic journals by using a five-point scale of 0-4 where 0 indicated not satisfied and 4 meant highly satisfied. Assigned scores for each item were converted into mean scores. Frequency of responses for each resource and respective means scores are given. It was found that elsevier's science direct was the most used resource by the faculty and students in all engineering disciplines and it is receiving the highest means core of 2.32. It was followed by IEL online was the next used resource by the faculty and students of computer, electrical and electronics, electronics and communication and it is receiving the mean score of 1.91. It was followed by .51 and .64 mean score of ASCE and ASME resources used by civil

and mechanical engineering students respectively. The other two resources of Nature & Nature biotechnology and management journals with a means score of .76 and .42 were mostly used by biotechnology and management students respectively.

Methodology:

The research scholars and faculties of various research organizations in Management Research Centers, Chennai, Tamil Nadu represented the target population for this study. The questionnaire method has been employed to collect the data for the present study. The questionnaire was constructed based on the following elements; use profile, frequency of visit, purpose of using online resources, acceptance of electronic media, awareness of the existence of electronic resources, publicity to promote the usage of electronic resources, paper presentation in conferences, papers published in Journals (National and International), publication details and reasons for non-use of electronic resources.

Data Analysis:

The data collected through questionnaires were organized and tabulated by using statistical methods and percentages. A total number of 450 questionnaires were randomly distributed among the faculty and researchers, out of which only 432 filled questionnaires were returned to the investigator. The response rate is 96 percent.

Sample:

The population of this study is faculty and Research Scholars in Management studies in Management Research centers, Chennai, Tamil Nadu. The sample consists of 432 faculty members and Ph D students who filled in the questionnaire of the annual user survey of 2009-2010. Thus, the sample consists of self-selected volunteers and may be biased towards those who are most active users of electronic resources. The analysis below indicates that the sample is reasonably representative by sex, discipline, occupation and university. However, some later findings hint that it is likely that active users of electronic resources are somewhat over-represented in the data. Inspection of the demographics of the respondents showed that the sample's breakdown by sex was nearly equivalent to the population. 50 % of university faculty and PhD students in Management Research Centers, Chennai, Tamil Nadu were women in 2009, whereas a in this study 50% of the respondent were women.

Suggestions:

On the basis of the response and opinion given by the respondents some of the important suggestions have been made, which will help the effective use of the electronic resources.

1. The authorities of the university and research institutions should take keen interest for providing better infrastructure facilities for the improvement of internet speed, so that users can feel more comfortable in browsing e-resources.
2. The university library and research institutions should make necessary arrangement for continuing subscription for the print journals along with e-journals. Since more than 95% of Research scholar are interested in making use of print journals.
3. The university library and research institutions should conduct orientation/training programmes regularly to assist the use of e-resources.

4. Majority of the research scholar have suggested that UGC-INFONET and other library consortium and they should provide pdf files of science direct, wiley-intercience and all other scientific journals.

Conclusion:

One of the major developments in libraries and information systems in the past 10 years is the advent and spread of electronic resources, information services and networks mainly as a result of developments in information and communication technologies. The change is basically of physical form where information content is increasingly being captured, processed, stored and disseminated in electronic form. The unique features of the information needs of users in electronic environments relate to the physical form in which information content is made available in electronic information environments. Users normally desire the content to be made available within the constraints of their skills and technological capabilities so that it is possible to access and use the required information content to resolve the felt gap in knowledge.

Reference:

- 1) Ikhizama, B.O & Oruwole, A.A. (2003) Pttren of usage of inforamton sources by scientist in Nigerian Universities of Agriculture (UoA), *Library Progress*, 23 (1), pp. 1-6.
- 2) Julien, Heidi, & Michels, David (2000) Source selection among information seekers: Ideals and realities, *Canadian Journal of Inforamation and Libray Science*, 25 (!), pp. 2-17.
- 3) Liu, Ziming (2006) , Print vs. Electroinc resources: A study of user perceptions, preferences and use, *Information processing and Management*, 42 (2), 583-592.
- 4) K.P.Singh and M.P.Satija, (2007) Information seeking behaviour of agricultural scientists with particular reference to their information seeking strategies: *Annals of Library and information Studies*. Vol.54, Dec 2007, pp.213-220.
- 5) M.Chandrashekara and K.R Mulla, (2007) The usage pattern of electronic information resources among the engineering research community in Karnataka: A survey. *Pearl Journal*; vol.1 No 4 oct-dec 2007. pp.33-38.
- 6) Alwarammal R, Sivaraj S and Madasamy R (2009), Promotion and usage of electronic resources by the students and members of faculty in engineering colleges in tamil nadu, india:An empirical study. *Knowledge networking in ICT era, International conference proceedings*, Vol 2. pp.676 - 679.

தொடக்கப் பள்ளிக்கு தேவையான

தமிழ்க் கணினி தேவைகள்

மா.ஆண்டோ பீட்டர்

தலைவர்-கணிததமிழ்ச்சங்கம்

செயற்குழு உறுப்பினர்-உத்தமம்

117.நெல்சன் மாணிக்கம் சாலை, சென்னை.29, இந்தியா

தொலைபேசி: +91-44-42113535

eMail: softviewindia@gmail.com | www.softview.in

கணினியையும் பள்ளி மாணவனையும் பிரிக்க முடியாத சூழல் ஏற்பட்டுள்ளது. பாடத்திட்டத்திலோ அல்லது பிரத்யேகமாகவோ அல்லது கூடுதல் சிறப்பு பாடமாகவோ கணினி படிப்பு வளர்ச்சி பெற்றுள்ளது. அயல்நாடுகளில் கணினியைக்கொண்டே அனைத்து பாடங்களையும் கற்க வேண்டிய சூழல்உள்ளது. இவை மொழி சார்ந்த பாடங்களுக்கும் பொருந்தும். ஆனால் தமிழகத்தில் ஆசிரியர்களும், பள்ளி மாணவர்களும் பெரும்பாலும் கணினியை இயந்திரமாகவே கருதுகின்றனர். இந்நிலை மாறி ஆசிரியர்களும், பள்ளி மாணவர்களும் தமிழ்மொழி வளர்ச்சிக் கான கருவியாக கணினியை கருதவேண்டும். மேலும் பள்ளி மாணவர்களுக்கு தமிழ்க்கணினி ஆற்றல் அவசியம் தேவை என்ற சூழல் ஏற்பட்டுள்ளது. மாணவனை தயார் செய்யும் ஆசிரியர் மற்றும் கல்வி மையமும் தகவல் தொழில்நுட்பத்தில் தமிழ்க்கணினி அறிவு பெற்றிருந்தாலே மேலோங்கி இருக்க முடியும்.

விண்டோஸ் இயங்குதளம், எழுத்துரு, எழுத்துரு வகைகள், எழுத்துருக்களை பொறுத்தும் முறை, எழுத்துக்கள் உள்ளீட்டு முறை, எழுத்துக்கள் உள்ளீட்டு வகைகள், அதன் சிக்கல்கள் மற்றும் குறைபாடுகள் எழுத்துரு வாரியாக நாம் அறிந்திருக்க வேண்டும். தமிழ் உள்ளீட்டு விசைமுறைகளாகிய தமிழ் 99, தமிழ் தட்டச்சு, தமிழ் ஒலியியல், தமிழ் பொனடிக் மற்றும் பிற விசை முறைகளை நாம் கற்றிருந்தலே விரைந்து பாடத்திட்ட பணிகளை கையாள முடியும். ஆசிரியர்கள் தொடக்கப்பள்ளி அளவிலேயே மாணவர்களுக்கு இவற்றை கற்றுக் கொடுத்தல் அவசியமாகும். மாணவர்களும் சுயமாக இது சார்ந்த பணிகளை தன்னந்தனியாக செய்ய துணிவுள்ளதா என ஆய்வுசெய்ய வேண்டும். இந்த தொழில்நுட்ப சார்ந்த வளர்ச்சிப்பணி களையும் உடனுக்குடன் மேம்படுத்தி தன்னை வளர்த்துக்கொள்ளவேண்டும். பெரும்பாலான இடங்களில் தமிழ்சார் வன்பொருள் மற்றும் மென்பொருளை வாங்குவதற்கே தயக்கம் உள்ளது. அடிப்படை கட்டுமானமாக மாணவர்கள், ஆசிரியர்கள் மற்றும் கல்வி மையம் எவ்வாறெல்லாம் மேம்பட்டுடன் இருக்க வேண்டுமென சிந்திக்கவேண்டும். தமிழில் விசைப்பலகைகள், சொற்செயலிகள் மற்றும் பல்வேறு மென்பொருட்கள் இருந்தாலும் வாங்கிப்பயன்படுத்து வதற்கு வெட்கமும், தயக்கமும் ஒரு காரணமாக உள்ளது. சுயமாக இவற்றை நாம் தவிர்த்து ஒவ்வொரு கல்வி மையமும் தமிழ் விசைப்பலகைகள், சொற்செயலிகள் மற்றும் பல்வேறு மென் பொருட்களை பயன்படுத்த

முன்வரவேண்டும். இவற்றின் மூலமாக நாம் கற்றல் தமிழ்க்கணினி பயன்பாட்டை சுலபமாக நாம் அறியமுடியும். நம் தமிழும் அறிவியல் மொழி தானே?

விக்கிபீடியாவின் பின்னணியாய் அமையும் கோட்பாடுகள், விக்கிபீடியாவின் வரலாறு, விக்கிபீடியா சமூகம் தொடர்பான அறிமுகம், கட்டற்ற திறந்த மூல இயக்கம், திறந்த புலமைச்சொத்து தொடர்பான அறிமுக விளக்கங்களும் அக்கோட்பாடுகள் விக்கிபீடியாவை எவ்வாறு உருவாக்க விழைந்தன. விக்கிபீடியா ஊடாக வெற்றியை நிரூபித்தன என்பனபோன்ற தகவல்கள் சிறுவயதிலேயே மாணவர்களுக்கு தெரியவேண்டும். அவ்வாறு போதிக்கப்பட்டால் தான் வருங்காலத்தில் தானும் விக்கிபீடியாவில் தமிழ்ப்பணியாற்ற ஆர்வம் பெருகும். தமிழ் வளர்க்கும் தமிழ் இணைய பல்கலைக்கழகம், அதன் தொகுப்புகள் மற்றும் டிஜிட்டல் டிலைப்ரரி ஆகியன இளம்வயதிலேயே மாணவர்கள் பயன்பெரும் வகையில் போதிக்கவேண்டும். தமிழ் இணைய ஆற்றல், மின்னஞ்சல் பயன்பாடு, தேடுப்பொறி பயன்பாடுகள், அனிமேஷன் வாயிலாக தமிழ்க்கணினி தயாரிப்பு, இணைய வாயிலான தேர்வு, தமிழ் மென்பொருட்கள் அறிவை வளர்த்தல் ஆகியவற்றை படிப்படியாக நாம் வளர்க்க வேண்டும்.

ஒரு மருத்துவரின் குறிப்பும், மருந்தும் பிறருக்கு புரியாததன் தன் காரணம் அவற்றிற்கான தமிழ் சொற்கள் இல்லாமையே ஆகும். தகவல் தொழில்நுட்பத்தை பொறுத்தவரை ஆயிரக்கணக்கில் தமிழ்மொழி சொற்கள் புழக்கத்தில் உள்ளது. குறிப்பிட்ட வட்டத்திற்குள்ளேயே புழக்கத்தி் லுள்ள இச்சொற்கள் வெளிக்கொண்டு வரவேண்டும். பள்ளிப் பருவத்திலேயே மாணவர்களுக்கு கலைச் சொற்களை பொருள்பட போதித்தால் நம் தமிழ்மொழி, மேலும் வளர்ச்சி பெறும். தொடக்கப்பள்ளி அளவிலேயே படத்துடன், பொருள்பட கணினி பயன்பாட்டை மாணவர் களுக்கு விளக்கி கற்பித்தால், மாணவ பருவத்திலேயே தமிழ் சொற்கள் ஆழமாக குழந்தைகள் மனதில் பதியும்.

ஆசிரியர்களும், பள்ளி மாணவர்களும் 2டி அனிமேஷன் அறிவை ஆழமாக பெறுதல் வேண்டும். 2டி அனிமேஷன் அறிவை தமிழுக்கு ஏற்றபடி பயன்படுத்துதல் வேண்டும். 2டி அனிமேஷன் அறிவை தமிழ் கற்பிக்கும் கருவி தயாரிக்கும் பணிக்கு ஏற்றபடி பயன்படுத்துதல் வேண்டும். 2டி அனிமேஷனில் கோரல்ட்ரா, போட்டோஷாப், பிளாஷ், டைரக்டர் ஆகிய மென்பொருட்களை கற்றிருத்தல் வேண்டும். இருபரிமாண உருவங்களை தயாரித்து தமிழ் கற்பித்தலுக்கு பயன்படுத்துதல் வேண்டும். தமிழ் உச்சரிப்புக்கும், பேச்சு தமிழுக்கும் கணினி மென்பொருட்களையும் அதன் கருவிகளையும் பயன்படுத்துதல் வேண்டும். 2டி அனிமேஷன் மூலம் பாடங்கள் தயாரிக்கும் அறிவை பெற்று வட்டாரச்சொற்களை வளர்க்கலாம். உதாரணத்திற்கு வீட்டை பெருக்கும் துடைப்பம் தமிழகத்தின் பல பகுதிகளில் வெவ்வேறு சொற்களாக அழைக்கப் படுகிறது. வாரியல், விளக்குமாறு, பெருக்குமாறு, தொடப்பம், வாருகோல் என பல வட்டாரச் சொற்களை நம் வசிக்கும் வட்டாரத்திற்கேற்ப பயன்படுத்தி மாணவர்களுக்கு பாடங்களை எளிமையாக்கலாம்.

தொடக்கப்பள்ளி மாணவர்கள் இங்கு குறிப்பிட்ட அனைத்து ஆற்றலையும் பெறமுடியுமா? சிறுபருவத்தில் அனைத்து சிறார்களும் தமிழ்க்கணினி அறிவை பெறமுடியும். தமிழ்க்கணினி அறிவை சிறுபருவத்தில் பெற்றால் தான் இளமைப்பருவத்தில் தமிழ்க்கணினி தயாரிப்பை பல கோணங்களில் நாட்டுக்காக படைக்கமுடியும். அயலக மென்பொருட்களுக்கு இணையாக, நாம் வடிவமைக்கும் முறை, கற்பிக்கும் முறை, இதன் பலன்கள் ஆகியன நம் தமிழ் வளர்ச்சி மற்றும் மேம்பாட்டிற்கு பாமரருக்கும் உறுதுணையாக இருக்கும்.

Quality Assessment Technique of E-Learning in Tamil Language

A. Kovalan

e-mail: kovapmu@gmail.com

Siva Pillai

e-mail: pillaisiva@gmail.com

Abstract

Web-Based Teaching Learning Process (WBTL) is a rapidly growing area in Education. Traditional forms of teacher education are transformed, as the Internet becomes a new medium for communication. Traditionally teachers have fulfilled dual roles as presenters of structured information and social agents in the educational process. Students are in need of good interactive resources with learning tools and techniques. Hence, there is a need for training in WBTL so as to enable the teacher to provide good resources in the web. The web-based learner resources can improve the quality of teacher education by availing various tools and techniques of assessment. The assessment of web-based learning resources helps to provide quality web resources in teacher education. It is also used to help teacher to have better resources and environment in which teaching takes place. The environment includes the organization, the learning materials, use of media, delivery methodology and various approaches in details. Assessment is a judgment regarding the worth or value of something. Typically the assessment process is divided into two parts. The first is a teacher assessment, which relates to interaction and guidance of a teacher with students and the second is a learning resources assessment, which relates to quality of materials and resources of a course. However, the primary function of the assessment is to help teacher to improve the total quality of education in web teaching learning environment.

1. Introduction:

The rise of e-learning and web-based training in Teaching Learning Process in Tamil Language (TLPTL) has lead to a growth in the use of web based learning assessment, which will increase as the use of e-learning becomes more widespread. Designing web-based tools, making resources and quality assessment of the learning resources is a challenging task for educators, curriculum designers, computer programmer and evaluator of the resources in the Web-Based Teaching Learning Process (WBTL). The teacher in online should take special effort for implementation and development of the instructional resources which is clear, accessible support, simplicity, transparency of the educational materials, and likelihood of the students' misconceptions as a result of interacting with materials. The teacher or trainer assessment is also very important to learners' motivation, self-regulated learning, continues education, knowledge improvement, people with disability and poor social background. This paper presents an overview of web based learning assessment cycle in two different aspect viz. Web Based Learning Teacher Assessment Cycle (WBLTAC) and Web Based Learning Resources Assessment Cycle (WBLRAC).

2. Web Based Learning Teacher Assessment Cycle:

An online teacher/trainer plays a key role in developing and maintaining an effective online learning environment and possesses a unique set of skills to success of a student. A quality of online teacher is very important than any resources in online. The role of online teacher is not just respect their need; they need to be involved in proper guidance, formative and summative assessment, counseling, administration and learners’ motivation. They should also support their feelings, and make facilitate to communicate with the learners with different media and methods. This means that the online teacher should make their time efficiently with their students. Hence, the quality of teacher/trainer can assess in six major categories viz.:

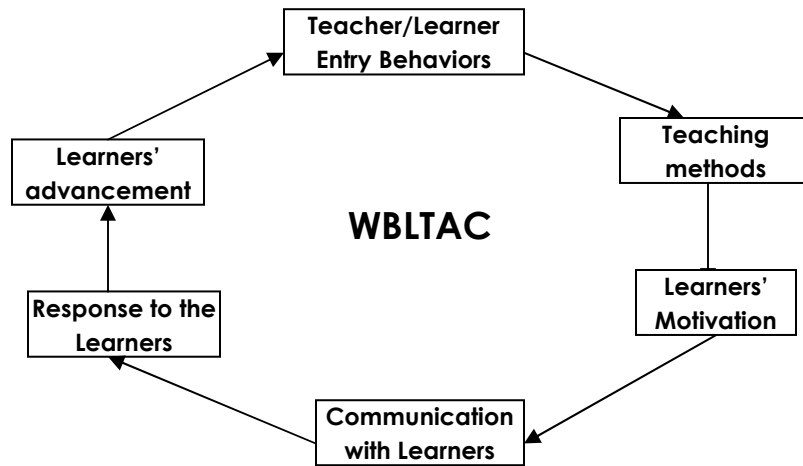


Fig. 1. Teaching Assessment Cycle (TAC)

2.1. Teacher/Trainer Entry Behaviors:

Once you have identified a person as an online teacher/trainer for a specific task or target groups the organizer determine what kind of knowledge, skill, and experience need to be taken into account for the specific training process. The entry behavior may be determined by three major areas viz.:

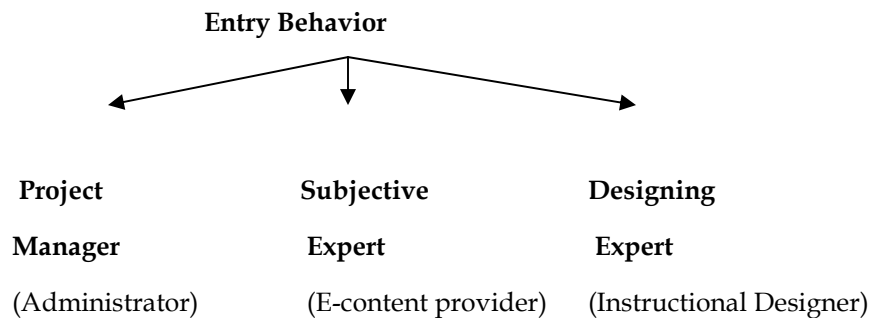


Fig. 1.1. Teacher Entry Behaviour

2.1.1. Project Manager:

Project manager is an achievement of academic qualification, administrative power, attitude level, communicative ability, leadership quality, involvement in students' achievement, investigative technique of students' problems; task oriented training activities, cooperation with student on assignment, equal treatment made to each student, responsibility, freedom, and support by giving opportunity to express their opinion and relationship with students, teacher and designer.

2.1.2. Subjective Expert:

Subjective Expert is based on theoretical and experimental experience in a different content area. The e-content provider should have knowledge of content development, material editing and method of teaching. E-content development is essential for learner, multimedia producer, Instructional designer and distributors. These are the quality needed for bearing on the delivering of the training product/event.

2.1.3. Designing Expert:

Designing expert should have knowledge of expectation of an online learner. The e-content material designer have different roles in a web based learning materials developing a project viz. instructional designer, multimedia designer, graphics artist, internet based application developer, media producer, and web administrator.

2.2. Teaching Methods:

Web based trainer use different methods to deliver an instructional material. The structure of an online teaching method consists of following three components.

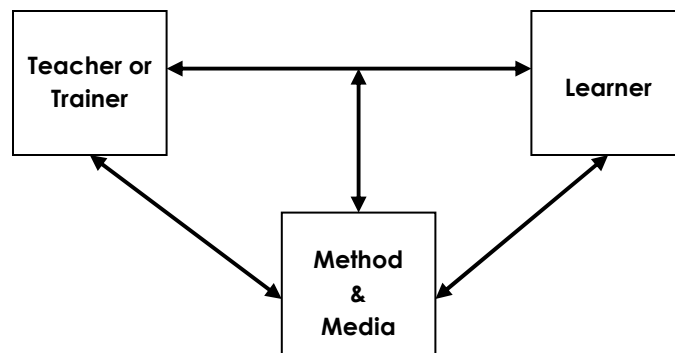


Fig. 1.2. Structure of teaching method

The trainer provides content in different media viz. text, pictures, audio, video, animation, and simulation which helps student to understand the concept easily. It implies that the learner has freedom to choose his choice by the way of his interest and preference of his learning materials.

2.3. Learners' Motivation:

Encourage learners to solve their problems by interaction with facilitator, peer group discussion, and expert in their chosen subject. The models of achievement motivation most often attribute students' academic motivation to cognitive processes (Bandura, 1986; Weiner, 1992) which regulate students'

learning behaviour. There is a growing body of evidence (Wentzel, 1991, 1994, 1996), however, that a consideration of social motivation of classrooms should not be excluded from the model of achievement motivation. Three aspect of learners' motivation is important in web based teaching process viz.:

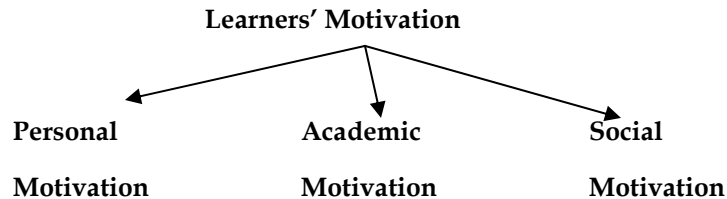


Fig. 1.3. Learners' Motivation

Success at online teaching learning process requires students to achieve three outcomes of education: individual power, academic achievement, and social adjustment. Judgments on levels of motivation were made using the following criteria: teacher interaction with learners and individual pupils, time keeping, teaching style, and observed enthusiasm in different aspect.

2.4. Communication with Learner:

Continues communication with learner helps to overcome the isolation problems. Web based teacher have many channel to communicate with their learner viz.

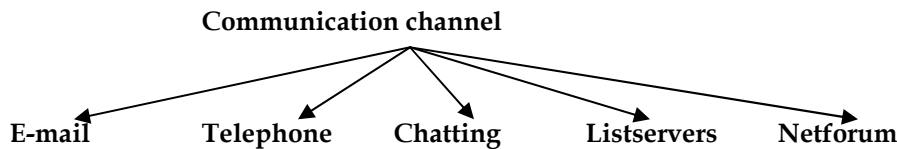


Fig. 1.4. Learners' communication channel

Electronic mail (E-mail) can be used by the facilitator and learners-to-learners communication. E-mail is cost-efficient, fast and convenient. Group e-mail can also be used to contact all learners simultaneously.

The telephone has used as a synchronous method of supporting the learners. The telephone allows learners to communicate with their teacher/trainer. Learners can discuss instructional and non-instructional problems with their facilitator.

Chatting helps to learner to communicate with his facilitator and his peer groups. Two types of chatting methods are available viz. video chatting and audio chatting. Video chatting is like a face-to-face communication and the audio chatting is like a telephone communication.

Listservers is an e-mail list that allows any of the users registered on the list to e-mail a specific listserv address, after which the e-mail is forwarded to everyone on the list. It is useful for the learner to discuss of topics relevant to the learning event.

Netforum allow the facilitator to post learning event information or announcements, changes to the learning event or deadlines for assignments. It is also used by learner to post questions about the learning event which the facilitator and other learner can answer. Learners and researchers are encouraged to discuss and reply in discussion forum.

2.5. Response to the Learner:

It is the quality of awareness that is evoked in collaborative meaning-making with students that defines the quality of a teacher's response to the teaching situation. Immediate feedback, which makes encourage, interest on further and continuing education, helps the learner to achieve their goal. A variety of possibilities exists. The teacher may have a word or phrase of affirmation:

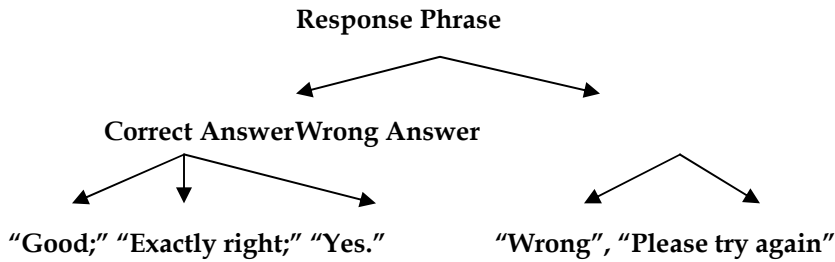


Fig. 1.5. Teachers' Response Phrase

2.6. Learners' Advancement:

Learners' record shows the quality of teacher/facilitator in web based teaching learning process. Ensuring the achievement of learners in different aspects viz. academic achievement, personal improvement, social behaviors, communication development, self-awareness, self-motivation, etc.

3. Web Based Learning Resources Assessment Cycle:

Instructional resources are plays an important role to deliver the learning materials through online. Using these resources, the learner can contact their facilitator at any time or in a fixed time when they are having problems. A quality resources make quality learning product and provide quality learner in web based training process. The quality resources can be assessed by six major categorize viz.

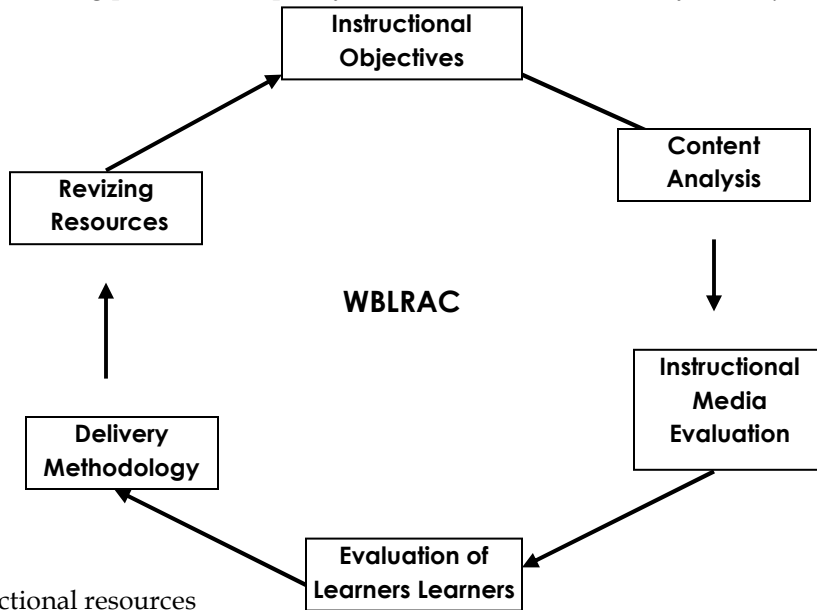


Fig. 2. Instructional resources

3.1. Instructional Objectives:

Objectives are useful in content development, material designing and implementation of the web based learning process and evaluation of the learner. The instructional objective describes a kind of performance expected by learners at the end of a learning event. The objective helps learner to plan the learning opportunity in their interested area. The development of instructional objectives is a task whose importance should not be overlooked. The instructional objectives can be divided into two major categories.

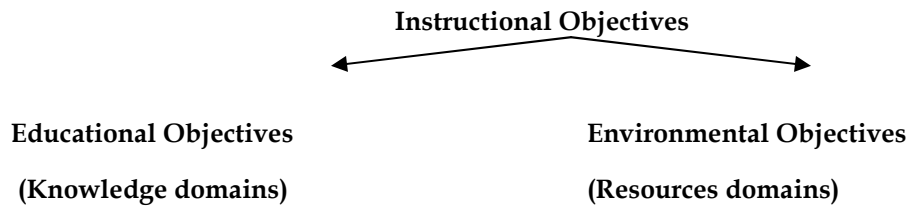


Fig 2.1. Learners' Instructional Objectives

3.1.1. Educational Objectives (EdO):

Educational objectives help learners know where they are going and how they are going to achieve their goal. EdO is designed based on learners' interest and involvement. In this domain the facilitator investigate the learners' interest to achieve their goal.

3.1.2. Environmental Objectives (EnO):

This environment helps learner to achieve their goal at their own location. In virtual learning environment the learner gets freedom to choose their interest subject with establish guidelines for including graphics, videos, audios, animations, pictures, and various presentation media. This environment makes more benefits to a learner to achieve their goal without any strain and pain.

3.2. Content Analysis:

The content review team might include the client, instructional designer, subject-expert(s) and programmer. Developing new Learning Materials (LM) is a major task involving their design, development, delivery and evaluation to ensure their effectiveness. Check the existing materials before providing a new one. If there is large number of subject matter and other information relevant to the learners' achievement goal on the web, then it is probably suitable for web-based delivery. The LM should provide more opportunity to the learner to choose their specific content area. The learning event needs to be updated regularly. In content analysis the following points need to consider for analyzing to provide a better learning materials.

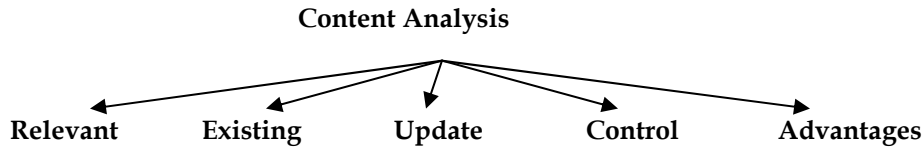


Fig 2.2. Learners' content analysis

Analyze, how well the content and material presented in the way that was both interesting and stimulating. The evaluation of these materials should include gathering data regarding the relevance of various assignments and the quality of the various assessments.

3.3. Instructional Media Evaluation:

In an online environment, media are used to provide variety to what is essentially a text-based methodology. Such media may include video, although this is limited unless learners have access to a broadband network, graphics, simulation, other visual effects and synchronous and asynchronous communication. Instructional media makes learner with self-learning, self-confident and self-control.

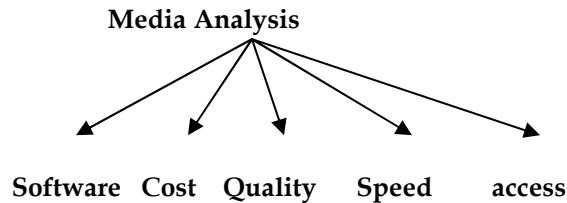


Fig. 2.3. Learners' Media Analysis

3.4. Evaluation of Learners:

This evaluation of reaction is based on measurement of learners' feelings and opinions about the course after completed. The facilitator can get the information here relates to methods of instruction, course content, learning methods and materials. Behavioral changes of learners are a measurement of the behavioral changes occurring as a result of the learning event after completed.

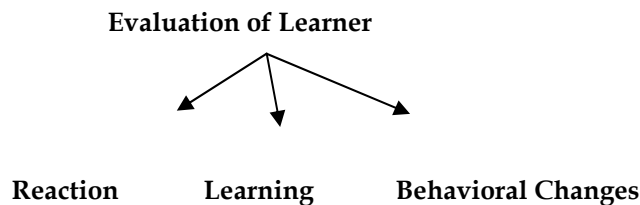


Fig. 2.4. Learners' Evaluation

3.5. Delivery Methodology:

The learning materials should be downloaded easily at learners place. The instructional media delivered with minimum space and maximum speed at learners' center. The material and media should flexible to the learners' environment.

3.6. Revising Resources:

The facilitator can determine the ability of the web system to support the learner, the effectiveness of the learning environment, the effectiveness of the support offered to the learner, how web based learning

benefits when compared with the traditional methods. Based on the above findings of the evaluation changes may be needed to the learning methods and materials to support the learner.

4. Conclusion:

Methods and media play a major role in the success of a web-based learning process. Hence we need to assess the material and media to provide a better learning event. A web-based learning process cannot be successful without well-designing learning material, good facility support and assessment. Web-based learning is a self-based leaning and places the responsibility for study directly on the learners themselves. They need to know how to use the web based learning environment to be successful.

References:

1. Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*, Englewood Cliffs, NJ; Prentice Hall.
2. Weiner, B. (1992). *Human motivation: Metaphors, theories and research*. Newbury Park, CA: Sage.
3. Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, 62, 1066-1078.
4. Wentzel, K. R. (1994). Relations of social goal pursuit to social acceptance, classroom behaviour and perceived social support. *Journal of Educational Psychology*, 86, 173-182.
5. Wentzel, K. R. (1996). Social goals and social relationships as motivators of school adjustment. In J. Juvonen, & K. R. Wentzel (Eds.), *Social motivation: understanding children's school adjustment* (pp. 226-247). New York: Cambridge University Press.

Impact of Multimedia technologies for Tamil education:-A global perspective

Prof. B. Jagadhesan

D.B.Jain College (Autonomous), Chennai-97
jagadhesanb@gmail.com Mobile:-9444532133

Introduction:

Tamil is a Dravidian language predominantly spoken in southern India and northeastern Sri Lanka, with smaller communities of speakers in many other countries. As one of the few living classical languages, Tamil has an unbroken literary tradition of over two millennia. The written language has changed little during this period, with the result that classical literature is as much a part of everyday Tamil as modern literature. Tamil school children, for example, are still taught the alphabet using the *átticúdi*, an alphabet rhyme written around the first century AD.

Modern technologies have started changing the lifestyle of our modern population. Multimedia, as a teaching aid, is very much effective with color, sound, graphics, which are found in the audio media and video media and movie media. One of the most significant changes in tamil education in recent years has been the availability of a range of Information Communication Technologies (Multimedia). Thus, Multimedia is no longer a new terminology nowadays. Almost everyone is familiar with Multimedia not only at work but also in schools as it encompasses the Internet and Multimedia.

The power of Multimedia can be used effectively in language teaching and learning as there is a paradigm shift from traditional teaching to using Multimedia in classrooms. It provides alternative teaching methods and it is a breakthrough from the traditional classroom environment. This is an efficient method to engage today's pupils. Multimedia tool is believed to provide the possibilities of multiple perspectives and a realistic learning environment. The real power of multimedia to improve education may only be realized when students actively use them as cognitive tool. Furthermore computer based learning is more motivating for students and this is generally accepted by educators and by administrators. Special innovative method can be supported to the young learners acquiring more vocabularies for suitable communication transactions in Tamil.

The contents have been presented with an elaborate use of multimedia technologies. The lessons of the educational programmes have been made more interactive and illustrative through pictures, animations, video clippings and audio contents. The library presents a picture-gallery and a video gallery on a large number of cultural events, temples and historical monuments. It also presents the devotional songs in audio mode. This paper presents the impact and experiences of developing multimedia based online resource for Tamil Education and culture history and art for global perspective.

Need of Multimedia in Teaching Tamil languages:

Explore the possibilities of the effective use of Multimedia in imparting Tamil language education. Analyze the potential of 2D and 3D software, audio, video, comics, comic strips and motion pictures in the acquisition of Tamil language skills. Identify the effective use of open source software and operating system for instructing Tamil language skills. Internet comprises innumerable features. It is the need of the hour to find out the means of using all those features for spreading the Tamil Language. Technological advancements in the field of 2D and 3D animations, Video, Audio, Comics, Comic strips, Motion films have made the Tamil teaching incredibly easy. Software and operating systems like Ubuntu(Tamil Linux),Microsoft, Microsoft Tamil office, suratha Unicode writer and converter, Google search engine and guruji search engine etc have greatly helped for the development of Tamil language. Internet acts as bridge for connecting all the media together. In today's scenario, web blogs and social net working websites work as major sources of alternative Mass media. Web blogs and social networking web sites play a significant role in imparting Tamil education all over the world in different forms-audio, video, 2D, 3D and motion pictures, etc.

Issues in Teaching Tamil:

The main issue with teaching Tamil is lack of quality teachers. Even though there is some limited support from government for community languages, the lack of demand from the Tamil speaking population has meant very few schools offering Tamil lessons.

As far as concerned there is no demand to employ a full time or part time Tamil teacher. The number of Tamil children in the school varies. if you look at the highly populated Asian community, there are community languages classes such as urdu, Punjabi, and Arabic in the mainstream school. This is not only because of the number of students. At present Tamil is at the same stage in some countries. Efforts at developing a syllabus and teaching framework are underway. As you will all agree, a strong fundamentals framework is a basis for a strong education system. This brings us to the role of multimedia in teaching and learning Tamil.

Real impact of Multimedia for teaching tamil language

The development of material that takes full advantage of computer-assisted learning used to be much more involved and typically employed interactive learning and Multimedia. A team comprising of interactive designers, graphic designers, programmers, testing team and a manager, usually develops such materials. A multimedia project may take up to six months of development time costing as much as \$70 000. Despite this, not many commercial products seem to have a substantial shelf life. Many outstanding titles done by big companies like Broderbund are seemed irrelevant (content-wise) in many countries. Like most teaching resources, multimedia packages vary greatly in complexity, quality and perceived usefulness. As with most educational materials, there are many approaches to teaching a topic, and even a well-produced program is unlikely to satisfy everyone. Therefore such commercial titles are difficult, if not impossible, to be updated modified or improved by content experts for their specific needs. For that, the original programmer must be available to access the source code. Since the expertise to use new teaching tools stays with the IT developers, rather than with the teachers, the typical teacher makes little or no progress in the use of interactive media tools. To date there are less than 20 quality Tamil language based educational CD ROMs available worldwide. Most of these titles are produced in

Malaysia. The content in these titles are based on the countries of their origin. Even though these titles score high in pedagogical value, the fact is there are still based on the local curriculum.

Authoring tools have improved rapidly. Some of these tools are easy to master. Time has come too drop the need for systems that make strong collaborations between IT Technicians and content experts (i.e. teachers). Now educationists can independently develop effective and customized instructional software for their own classrooms. In the old days of programming, the developer had to do everything in such arcane general purpose compiler languages like Assembler, Basic or C. However, now, many sophisticated resources are available readily. These include vast libraries of media, which can be copied and pasted in a program. Such clip-arts and sound clips are available on the Internet and CD ROM packages. There is also a range of tools available for each medium. From high-end professional packages to less sophisticated but still powerful and useful freeware offerings are available for the initiative programmer. The advent of easy-to-use multimedia authoring packages (like Mediator), and the widespread availability of computers, and the ease with which traditional classroom teaching can be put into instructional software have all created an environment in which, traditional classroom teaching can be transformed and made more exciting for students and teachers.

Conclusion:

The perennial problem faced by emigrant Tamil communities is getting quality resources to teach Tamil to the younger generation. Multimedia development has actually become quite easy. I hope that this paper would inspire the audience to get together and develop their own multimedia teaching and learning pedagogical materials.

References:

1. Karrer, T Understanding eLearning 2.0
2. <http://www.learningcircuits.org/2007/0707karrer.html>
3. Nichols, M. E-Learning in context.
4. [http://akoaootearoa.ac.nz/sites/default/files/ng/group-](http://akoaootearoa.ac.nz/sites/default/files/ng/group-661/n877-1---e-learning-in-context.pdf)
5. [661/n877-1---e-learning-in-context.pdf](http://akoaootearoa.ac.nz/sites/default/files/ng/group-661/n877-1---e-learning-in-context.pdf)
6. Turgut, Y., & Irgin, Pelin (2009). Young learners' language games learning via computer games. *Procedia Social and Behavioral Sciences* 1 (2009) pg. 760-764 Vaughan, Taay. (1997). *Multimedia making it works*. (3rd edition) New Delhi: Tata McGraw Hill.

இன்றைய மாணவர்களுக்கான தமிழ் கணிமை

திட்டங்கள் – எங்கள் கல்லூரி அனுபவம்

முனைவர் ஆ. முத்துக்குமார், பேராசிரியர்,
குமரகுரு பொறியியல் கல்லூரி, கோவை, இந்தியா.
தொலைபேசி – 9489608402,
மின் அஞ்சல்: aru.muthukumar@gmail.com

கட்டுரை சுருக்கம்

தமிழ் கணிமை திட்டங்கள் என்றாலே ஒரு வித தயக்கம் இன்றைய கல்லூரி மாணவர்கள், ஏன் ஆசிரியர் களிடமும் இருக்கிறது என்பது கசப்பான உண்மை. இது இக்கால சூழ்நிலை காரணமா அல்லது படித்தவர்களிடம் தமிழ் பற்று குறைவா என ஆராய்வது ஒரு கோனம். அதை விடுத்து, அத்தகைவரிடம் நம்பிக்கை கொண்டு வர தமிழ் ஆர்வலர்கள் என்ன செய்ய வேண்டும் என ஆராய்வது இக் கட்டுரையின் நோக்கமாகும். எங்கள் கல்லூரியில் தற்போது மூன்று மாணவர்கள் திட்ட நிறுவலை மேற்கொண்டு வருகிறார்கள். நிறைய மாணவர்கள் முதலில் ஆர்வம் காட்டினாலும், பின்னர் தங்கள் கருத்தை மாற்றி கொண்டு விட்டனர். இதன் காரணங்களை ஆராய்வதுவும் இக் கட்டுரையின் நோக்கமாகும். மாணவர் களிடம் நம்பிக்கை ஏற்படுத்தும் முயற்சியில், இக் கட்டுரை ஆசிரியர், தகவல் தரவு தானியங்கி ஆராய்வு மென்பொருளை உருவாக்கும் முயற்சியில் ஈடுபட்டுள்ளார்.

1. முகவுரை:

தமிழ் கணிமை என்றாலே ஏதோ ஒரு வலை பூ நிறுவுதல் என்ற எண்ணம் அநேக நபர்களிடம் உள்ளது. இன்னும் பலர், தமிழில் எதற்கு, இதனால் என்ன பலன் என்று தட்டி கழிக்க பார்க்கின்றனர். தமிழில் மென்பொருள்களை உருவாக்கும் தேவையை நிறைய பேர் உரை மறுக்கின்றனர். மாணவர்களிடமும் இந்த குழப்பம் உள்ளது. இதை நீக்க நாம் தமிழில் பல மென்பொருள்களை உருவாக்க வேண்டும். இதை யார் செய்வது? இன்றைய மென்பொருள் வல்லுனர்கள் தங்களின் பொழுது போக்காக இதை செய்யலாம் அல்லது இதை ஒரு வியாபார நோக்குடனும் செய்யலாம். தமிழ் இன்று ஒரு நாட்டுக்கு மட்டும் சொந்தமான மொழியன்று. அநேக நாடுகளில் தமிழர்கள் வாழ்கின்றனர். அவர்களுக்கு உதவும் பொருட்டாகவும் இதை உருவாக்கலாம். தமிழ் இனி மெல்ல சாகும் என கூறுபவர்களிடம் நாம் இந்த மென்பொருட்கள் மூலம், தமிழ் இனி வேகமாக வளரும் என கூற வேண்டும். இந்த மென்பொருட்களை உருவாக்க மாணவர்களை பயன்படுத்தலாம். அவர்களின் குழப்பத்தை நீக்கி, நம்பிக்கையை ஏற்படுத்தும் பொறுப்பு, தமிழ் ஆர்வலர்களிடமும், ஆசிரியர்களிடமும் உள்ளது. இதற்கு உத்தமம் போன்ற அமைப்பினர், நிறைய பட்டறைகளை கல்லூரிகளில் நடத்த வேண்டும். பல்லூடகங்களில் நிறைய கட்டுரைகளை எழுத வேண்டும். தமிழ் கணிமை அனைத்து மக்களையும் சென்றடையுமாறு பார்த்து கொள்ள வேண்டும். இது ஏதோ ஒரு அமைப்பு அல்லது ஒரு சில நபர்களின் பொறுப்பு என பாராமல், அனைத்து ஆர்வலர்களும் ஊர் கூடி தேர் இழுக்க முயற்சிக்க வேண்டும்.

2.எங்கள் கல்லூரி அனுபவம்:

இந்த கட்டுரை ஆசிரியர் 2009 தமிழ் இணைய மாநாடில் கூறியவாறு, 12 மாணாக்கர்களை வைத்து தமிழ் கணிமை திட்ட நிறுவலை தனது கல்லூரியில், தொடங்கினார். ஆனால், மாணவர்களிடம் உள்ள தயக்கத்தை அவரால் நீக்க முடியவில்லை. பல்வேறு காரணங்களுக்காக 9 மாணவர்கள் வேறு திட்டங்களுக்கு மாறி விட்டனர். முக்கியமான காரணம், வேலை வாய்ப்பு பற்றியது. தமிழ் கணிமையில் திட்டம் செய்தால், பிறகு வேலை கிடைக்காது என்ற அவநம்பிக்கை ஆகும். இதற்கு ஆசிரியரால் ஒன்றும் செய்ய இயலவில்லை. மாணவர்களுக்கு நம்பிக்கை ஊட்டும் விதமாக பேசி பார்த்தார். பலனில்லை. அடுத்த காரணம், தமிழ் கணிமையில் என்ன செய்யலாம், என்ன செய்யமுடியாது என ஆசிரியரால் தீர்மானமாக கூற இயலவில்லை. இதற்கு, ஆசிரியரின் தமிழ் கணிமை பற்றிய அறிவு அப்போது குறைவாக இருந்ததும் ஒரு காரணம். அப்போது ஆசிரியரும் ஒரு கற்றுக்கொள்ளும் நிலையில் இருந்ததும் காரணமாகும். எனவே மூன்று மாணவர்களுடன் தொடங்கிய திட்டங்கள் இன்று நிறைவு நிலையை எட்டியுள்ளது. இந்த திட்டங்கள் பின்வருமாறு:

1. தமிழ் மின் கற்றல் – குழந்தைகளுக்கானது.
2. அகராதி நிறுவுதல் மற்றும் வலை தளம் நிறுவுதல்.
3. தமிழ் பேசும் மென்பொருள்.

எங்கள் அனுபவம், தமிழ் தட்டச்சு கற்றலில் தொங்கி, தமிழ் தரவு தளத்தை நிறுவுதல் மற்றும் தமிழில் மென்பொருள் எவ்வாறு நிறுவுவது என்பது வரை நடந்தது. நாங்கள் உபயோகித்த மென்பொருள் நிறுவும் கருவிகளானது பின்வருமாறு:

1. PHP, MySQL
2. Java, MySQL.
3. Flash

ஆசிரியரும் மாணவர்களுக்கு உதவும் பொருட்டு தானே மென்பொருள் சோதனை மேற்கொண்டார். ஆசிரியப்பணிக்கு இடையில், நேரம் கிடைக்கும் போது அவர் இதை மேற்கொண்டார். மாணவர்களுக்கு முழு நேர திட்டப்பணி ஆனாலும், அவர்கள், இரு வாரத்திற்கு ஒரு முறையே ஆசிரியரை சந்தித்து ஆலோசனை மேற்கொண்டனர். இந்த முதல் முயற்சி எதிர்பார்த்த வெற்றியை கொடுக்காமல் போனாலும், நிச்சமாக இது தோல்வியல்ல. இந்த அனுபவம், இக் கட்டுரை ஆசிரியரை, மேலும் இரு மிகப் பெரிய திட்டங்களை தானே எடுத்து செய்ய தூண்டியது என்றால் மிகையாகாது.

3. திட்டங்களின் நோக்கங்கள்:

அ) தமிழ் மின் கற்றல் – குழந்தைகளுக்கானது

இது 7 வயதுக்கு குறைவான குழந்தைகள் எளிதில் தமிழ் கற்க உதவும். இதில், எழுத்து, எண்கள், வார்த்தைகள், ஒலி அமைப்பு போன்றவைகளை ஒரு எளிதான முறையில் பயிற்றுவிக்கப்படும். இதற்கு, முதுநிலை கணினி பயன்பாட்டு துறை மாணவி திட்ட நிறுவலை மேற்கொண்டுள்ளார். இப்பணி, சூன் மாதம் நிறைவடையும் என எதிர்பார்க்கப்படுகிறது.

ஆ) அகராதி நிறுவுதல் மற்றும் வலை தளம் நிறுவுதல்

இந்த திட்டமானது, இரு நோக்கங்களை கொண்டது. முதலாவது, ஒரு தமிழ் வலை தளத்தை நிறுவுவது. இதில், மற்ற தளங்களில் உள்ளது போல அனைத்து தகவல்களைப் பெறலாம். செய்தி, விளையாட்டு, மாணவர்கள் பக்கம், பெண்கள் பகுதி, மருத்துவப் பகுதி போன்றவைகள் இடம் பெறுகின்றன. இரண்டாவது நோக்கமானது தமிழ் அகராதி பற்றியது. இதில், ஆங்கிலம் -> தமிழ், தமிழ் -> ஆங்கிலம், தமிழ் -> தமிழ், பேரகராதி போன்றவைகள் இடம் பெறுகின்றன. இத்தகைய மென்பொருள்கள் ஏற்கனவே

இருந்தாலும், எங்கள் அனுபவத்திற்க்காக சோதனை முயற்சியாகவே இந்த திட்டத்தை செயல்படுத்தியுள்ளோம்.

இ) தமிழ் பேசும் மென்பொருள்.

இந்த திட்டமானது, முழுதும் நிறைவடையவில்லை. ஒரு சோதனை முயற்சியாக மேற்கொண்ட இந்த திட்டமானது, 20 சதவீதமே நிறைவடைந்துள்ளது. இந்த திட்டம் தொடர்ச்சியாக வேறு சில மாணவர்களால், எதிர்காலத்தில் மேற்கொள்ளப்படும்.

இந்த மூன்று திட்டங்களுமே மாணவிகளால், ஆர்வத்துடன் மேற்கொள்ளப்பட்டுள்ளன. இக் கட்டுரை ஆசிரியரும் தானே சிறு, சிறு மென்பொருள் சோதனைகளை மேற்கொண்டு, மாணவிகளுக்கு வழி காட்டியுள்ளார். இந்த அனுபவமானது மேலும், சில திட்டங்களை எடுத்து செயல்படுத்த இக் கட்டுரை ஆசிரியரை தூண்டியுள்ளது என்றால் மிகையாகாது. உத்தம உறுப்பினர்கள் பலரும், ஆசிரியருக்கு, திட்ட நிறுவலுக்காக உதவியுள்ளனர். பல மின் குழும உறுப்பினர்களுக்கு, இக் கட்டுரை மூலம் நன்றி செலுத்த இக் கட்டுரை ஆசிரியர் கடமைப்பட்டுள்ளார்.

4.எதிர்கால திட்டங்கள்:

பின்வரும் எதிர்கால திட்டங்களை இக்கட்டுரை ஆசிரியர் மேற்கொண்டுள்ளார்.

1. Data Mining Tamil Web Pages or Text Files.
2. E-learning package for teaching Tamil Grammar using Tholkappiam.
3. Translation Software (English to Tamil).

இதில், முதலாவது திட்டமானது, தற்போது சோதனை செய்யப்பட்டு வருகிறது. சோதனை இதுகாறும், வெற்றியாகும். இக்கட்டுரை ஆசிரியரின், முனைவர் படிப்பில் ஆங்கிலத்தில் இத்திட்டத்தை வெற்றிகரமாக செய்து முடித்து, முனைவர் பட்டமும் பெற்றுள்ளார். தமிழில் இதை எளிதாக செயல்படுத்த முடியும்.

மூன்றாவது திட்டத்தை செயல்படுத்த தொல்காப்பிய இலக்கண வரை முறைகளை கற்க வேண்டியுள்ளது. இதன் முயற்சியாகவே, இரண்டாவது மின் கற்றல் திட்டத்தை செயல்படுத்த உள்ளோம். திட்டப்பணி, இரண்டாவது முடிவுக்கு வரும் தறுவாயில், திட்டம் மூன்றை எடுத்து செயல்படுத்த உள்ளோம்.

தொல்காப்பியம் மின் கற்றலில் கொண்டு வந்தால், தமிழ் இளநிலை, முது நிலை மாணவர்களுக்கு மிக உதவியாக இருக்கும். தமிழும் காலத்திற்க்கேற்ப தகவல் தொழில்நுட்ப வரையறைகளுக்குள் வர இத்தகைய மென்பொருள்கள் உதவும்.

மாணவர்களுடன் சேர்ந்தோ அல்லது தனி முயற்சியாகவோ இத்திட்டங்களை இக்கட்டுரை ஆசிரியர் மேற்கொண்டுள்ளார். முதலில், செயல்படுத்தினால், மாணவர்களுக்கு ஆர்வம் வரும். பின்னர் மாணவர்களை கொண்டு, மிக வேகமாக திட்டங்களை செயல்படுத்தலாம்.

5. முடிவுரை.

எங்கள் கல்லூரி அனுபவத்தை எடுத்து கூறுவதால், இது போல ஏனைய அனைத்து கல்லூரிகளும், முதலில் ஆசிரியர்களுக்கு பட்டறைகள் மூலம் கற்பித்து, எங்களைப் போல, மாதிரி திட்டங்களை எடுத்து செயல்படுத்தி மாணவர்களிடம் நம்பிக்கையை வரவைக்க வேண்டும். மாணவர்கள் மிகப் பெரிய சக்தி. அது இலவசமாக கிடைக்கிறது. மேலும், மாணவர்களுக்கு தமிழின் பால் ஆர்வம் வந்து, செயல்பட்டார்கள் என்றால், நாட்டுக்கும் நன்மை பயக்கும். தமிழ் மென்பொருட்களால், தமிழ் தெரிந்த மக்களுக்கு மிகவும் பயனுள்ளதாகவும், அவர்கள் அதை தங்கள் வாழ்க்கை தரத்தை மேம்படுத்த செய்வர் என்பது திண்ணம்.

Teaching Primary Education in Tamil Using LMS and Visualization Techniques

Richards Hadlee

Final Year M.Sc Computer Science,
Loyola College, Chennai
hadleefdo@gmail.com

R. E. Iniya Nehru,

Technical Director, NIC
Chennai
nehru@nic.in

Abstract

Extend teaching primary education in Tamil through Visualization technique. Visualization is creating images, diagrams, or animations to communicate a message. It's an effective way to communicate both abstract and concrete ideas. Visual representation of information requires merging of data, computer graphics, design, and imagination. Moodle provides only text information (report, items analysis, etc) and does not provide visualization tools. We use third-party tools to do visualization for LMS. Visualization helps to communicate results and understand better.

Keywords: Moodle, LMS, Visualization

1. Introduction & Purpose

LMS (Learning Management System) is highly interactive teaching tool. It offers great variety of workspaces to facilitate information sharing and communication among participants. It also provides a platform for a teacher to create/deliver contents, monitor student participation, and assess student performance through web application; this helps wider audience of students across the world (Education-Anytime and Anywhere).

E-learning classes most of the time is asynchronous. In this mode the instructor and students can interact via messaging or e-mail, and assignments can also submitted. The students and the instructor can be online at the same time to communicate directly and share information if required. E- Learning can include training, delivery of just-in-time information and guidance from Teacher.

A LMS (Learning Management System) is used to organize an online learning environment. The term "online" refers to Internet and Intranet environments. It can include images, text, audio, video, animation and virtual environments.

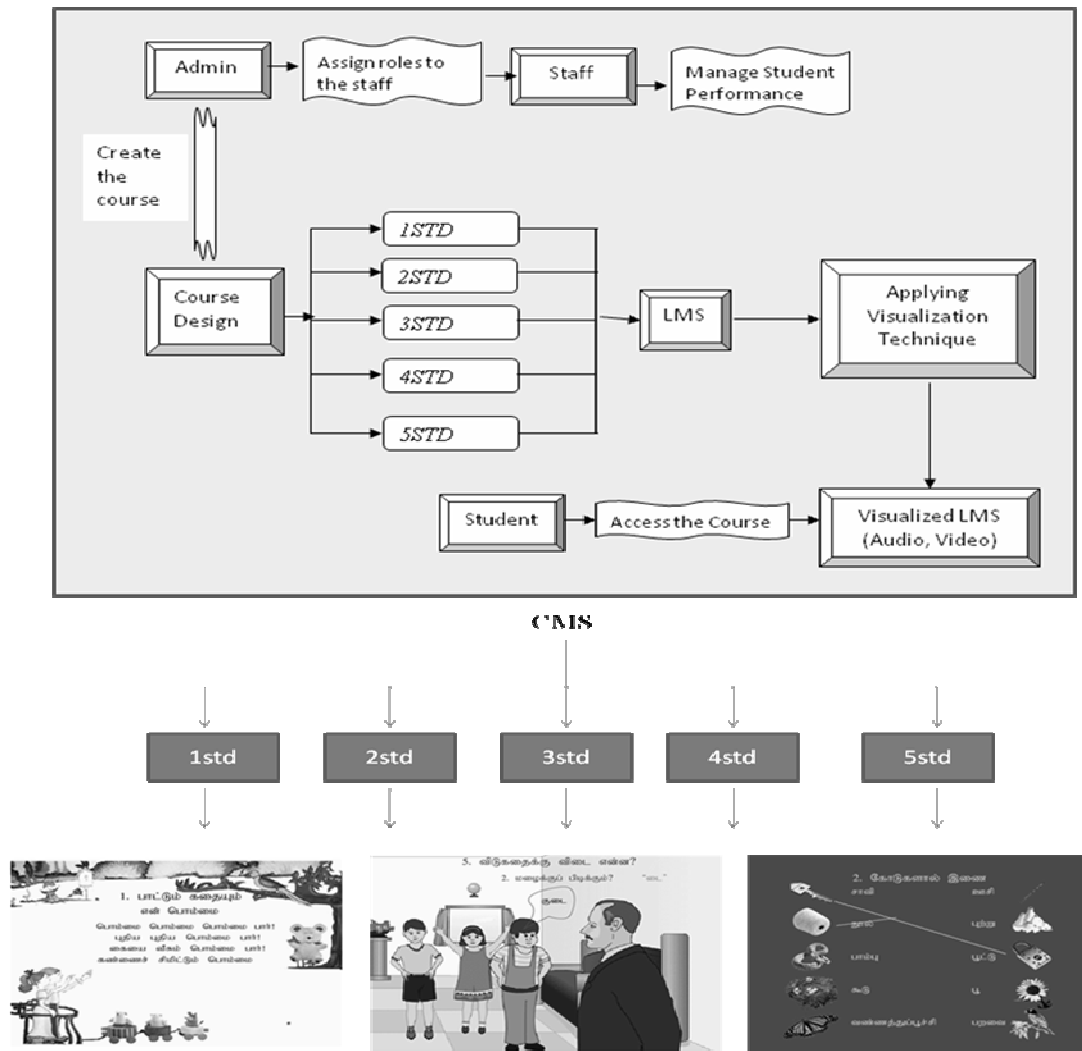
LMS and Visualization Techniques help to teach the primary school subjects by representing interactive or animated flash image files. The child chooses the learning style which has the seeing, feeling. Understanding learning styles is only a first step in maximizing potential and overcoming learning differences.

1.1 Lists of Functionalities

- Creating LMS for Primary education
- Applying visualization
- Questions and log reports

1.1.1 Creating LMS for Primary education

This LMS has been designed for primary education (1st to 5th Standard) in Tamil language. Each standard has its own subjects (Tamil, Mathematics, Science, and Social Science). These contents have been designed using visualization techniques.



1.1.2 Applying visualization

Visualization is a branch of computer graphics which is concerned with interactive presentation / animated digital images for users to grasp easily. These techniques facilitate analysis of large information by representing visual display. The instructor and students can interact with forum, chat &etc.

1.1.3 Questions and Reports

Each lesson contains the various types of exercises, through which the user can perform exercises in an effective way.

The grades are provided for the particular question which is performed by the user. The status of the question also is displayed at the end of the test.

If the user answered correctly, right status is displayed and grade has been marked as 1/1. If the user selected the wrong answer, no grade is provided.

Reports

The Reports is represented through a visualized pattern (Graph). This contains the number of days the student has worked on this. In this graph the x-axis consists the date and month he worked in and the y-axis consists of the levels of work he has done on the particular day.

Log report

Log Reports represents the lists of user who enroll into the particular course. Through this log the admin can get the performance of the user.

Activity Report

An activity report has the number of activity in that particular course. It contains the number of available resources for the course.

Participants Report

The Participants reports consists of the following criteria,

- What Activity?
- At what day?
- Reports of admin or user?
- What actions?



2. Implementation of Tamil Learning in Moodle

This application has designed and developed through Moodle, PHP & Mysql using Visualization Techniques. The purpose of this application is to bring awareness about e-learning using Moodle. Moodle is the easiest and most flexible LMS. It's easy to navigate, had features that are directly applicable to the writing classroom and best of all, is free for me to download and customize. It has a strong support community and good online documentation to help us. Moodle is specifically designed with educators in mind, allowing for easy setup and maintenance.

Primary Education

Primary School Teaching is a social networking and resource sharing site made exclusively for Teachers. It also allows other members of staff working in Primary Schools to participate. It is a platform for sharing teaching resources and ideas. It allows teachers to communicate effectively in a collaborative environment. Proof of that is, you will see many times a kid playing with his toys and imagine that he is in a battle field or a camp and they really feel it. All the subjects can be created with questions, so that the users get grade for any subject.

It is a process of transforming information into a visual form that enables user to observe the information. On the other side, it uses techniques of computer graphics and imaging. Successful visualizations can reduce the time to get the information, make sense, and enhance creative thinking. We can use visualization for improving memory, restoring health, reducing stress, increasing relaxation.

The following properties are used in visualization

- Text
- Graphics
- Images
- Video
- Audio
- Animation

3. Approach and Design

LMS and Visualization Techniques help to teach the primary school subjects by representing interactive or animated flash image files. Understanding learning styles is only a first step in e-learning that overcome the learning differences. There is no doubt of the effectiveness of visualization in every area of life, even a child does visualization although they don't know what it is.

Resources

The course has the lessons in it. The lists of resources are available in each lesson in the particular course. The lessons have been implemented in the tool called eXe. The eXe project developed a freely available Open Source authoring application to assist teachers and academics in the publishing of web content without the need to become proficient in HTML or XML markup.

Resources authored in eXe can be exported in IMS Content Package, SCORM 1.2, or IMS Common Cartridge formats or as simple self-contained web pages.

The content material in the particular lesson also been designed in Flash, in which a students can be attracted to learn difficult subjects through interactive audio & video based content.

The LMS has also been implemented using Visualization and Computer graphics in order to improve the courses on the students learning in Tamil.



Three main uses of visualization are:

- **Motivation.** Creative visualization is a great way to see a possible future and move you towards it.
- **Mental practice or rehearsal.** Mental practice or mental rehearsal is complementary to real practice. Mental practice can also be cost-effective and safer.
- **Reinforcing other techniques.** Visualization is a powerful way to strengthen other techniques, such as association and scripting

5. Conclusion

In this work we have shown application of teaching education in Tamil using LMS along with a visualization technique. We have described how different visualization techniques can be used in order to improve the courses on the students learning in Tamil. All these technique affiliated to the student's performance.

The animated image avoids the monotonous feeling of reader which we normally get in class room. Presently most of the children are computer friendly, spend lot of time in computer games, animation etc. Hence developing a course content using LMS tools like Moodle, PHP & MYSQL gives a different visual presentation to children. The students can be attracted to learn difficult subjects using Visualization and Computer graphics. The retention rate for an average student in found to be much more using LMS techniques compared to traditional classroom training.

Acknowledgment

I express my sincere thanks to **Mr. E. Iniya Nehru**, Technical Director, NIC (Chennai) His encouragement and enthusiastic support throughout this paper induced me to do well. I also thank **Mr. Jeyakumar S**, the Project Manager at NIC who helped me to get the needed information and encouraged for successful completion of this paper.

6. References

Books:

- Using Moodle 2nd Edition,
By Jason Cole and Helen Foster
Published by O'Reilly Media, Inc, 1005 Grayenstein Highway North,
Sebastopol, CA 95472

Articles:

- Factors in the deployment of a Learning management system at
The University of the South Pacific
Moodle For Teachers, Trainers And Administrators
Moodle An electronic classroom

Resources:

- <http://download.moodle.org/>
- <http://exelearning.org/wiki>
- www.docebo.org

கலைஞர் தமிழ்பேசும் கணினிக்கண்

உருவாக்கத் திட்டம்

முனைவர் அர. ஜெயச்சந்திரன்,

பேராசிரியர் மற்றும் தலைவர்,

பார்வையற்றோருக்கான மென்மங்களைத் தயாரிக்கும்

பணிக்கான தமிழியல் ஆய்வு மையம்,

பாரதிதாசன் பல்கலைக்கழகம், திருச்சிராப்பள்ளி - 620 024.

மின்னஞ்சல்: drj3363@yahoo.co.in கைப்பேசி : 9444337980

முன்னுரை

கண்பார்வையற்றோரின் பார்வையின்மை என்னும் துன்பத்தை நீக்கிடக் கணித்தமிழ் வழியாக மேற்கொள்ள வேண்டிய பணிகளை இக்கட்டுரை விரிவாக விளக்கிப் பேசவுள்ளது.

1970-களில் அன்றும் முதலமைச்சராக இருந்தவரும் இன்றும் உள்ளவருமாகிய மாண்புமிகு முதலமைச்சர் கலைஞர் அவர்கள் முதன்முதலாக, தாமாக முன்வந்து பார்வையற்றோருக்கு இலவசப் பயணச் சலுகையை வழங்கினார்; அதைத் தொடர்ந்து அவரது ஆட்சிக் காலங்களில் கண்ணாடி இழந்தோருக்குக் கண்ணாடி வழங்கும் திட்டம், பார்வையற்றோருக்குக் கல்வி மற்றும் வேலைவாய்ப்பு வழங்குதல், உயர்கல்வி நிறுவனங்களில் உயர்பணி வாய்ப்புகள் வழங்குதல் முதலிய பேருதவிகளைச் செய்து வருகிறார்.

இப்போதும் பார்வையற்றோர் உள்ளிட்ட மாற்றுத் திறனாளிகள் அனைவரும் ஏற்றம் கண்டிடத் தனித்துறை அமைத்துள்ள தமிழினத் தலைவராய் விளங்குகின்றார். இன்று தமிழ்பயின்று பணிசெய்யும் பார்வையற்றோர் கலைஞரின் தமிழ்க்குரல் கேட்டு, தமிழ்பயின்று பணிசெய்தல், சிறப்பாகச் சொற் பொழிவு நிகழ்த்தல், நூல் எழுதுதல் எனும் தமிழ்த்திறன்களைப் பெற்றுள்ளனர். இந்நிலையில் தமிழில் பேச்சுப்பொறி வடிவமைத்தல், அதனைக்கொண்டு பேசும் கணித்தமிழ்க்கண் உருவாக்குதல் ஆகிய வற்றில் கலைஞரின் குரலை இணைத்தல் மூலம் வாழ்நாள் முழுவதும் அவரே இருந்து விழியற்றோருக்கு வழிகாட்டுவார்.

இதனைத் தொடர்ந்து அவர் பெயரிலும் அவர் குரலிலும் தமிழில் பேசவல்ல கணினிக்கண் உருவாக்கும் திட்டத்தினை உருவாக்கிச் செயல்படுத்துவதன் மூலம் பார்வையற்றோர் அனைவரும் எழுதுதல், படித்தல், சுற்றுப்புறத்தில் உள்ள அனைத்துப் பொருள்கள் பற்றிய தகவல்களை அறிதல், சாலைத் தடங்களை அறிந்து தாமே இயங்குதல், வாகனங்களைச் செலுத்துதல் முதலிய அனைத்தையும் கணினிக் கண் கொண்டு செய்ய முடியும் எனும் கருத்தினை நிறுவுதலே இக்கட்டுரையின் நோக்கமாகும்.

மேல்நாட்டார் ஆங்கில வழியிலும், உருசியர், சீனர், ஜப்பானியர் போன்றோரும் கணினிக்கண் பற்றி ஆய்வுகளை மேற்கொண்டு படிப்படியாக மெய்ப்பித்து வருகின்றனர். தொடக்க நிலையில் அவர்களுடன் இணைந்தும் பின்னர் நமது அறிவைக் கொண்டும் தனித்தமிழில் கணித்தமிழ் வழியே இப்பணிகளைச் செய்திடும் நோக்கில் பின்வரும் அடிப்படைகளைக் கொண்டு இக்கட்டுரை அமைகிறது.

2006-ஆம் ஆண்டுச் சூலைத்திங்கள் மலையாளில் நடைபெற்றப் பன்னாட்டுக் கல்விக் குழுமத்தின் 12-ஆவது மாநாட்டில் "கருணைமிகுத் துணைமைத் தொழில் நுட்பம்" எனும் தலைப்பில் இக்கட்டுரை யாளரால் அங்கு ஆங்கிலத்தில் ஆராய்ச்சிவுரை நிகழ்த்தப்பட்டது. அதன் இறுதிப் பகுதியில் செயற்கைக் கண்பார்வை உருவாக்கம் பற்றிப் பேசப்பட்டுள்ளது. அதன் விரிவே இதன் தனிக்கட்டுரையாகும்.

அறிவுப்பார்வை

“அறிவு அற்றம் காக்கும் கருவி” என்று வள்ளுவர் கூறினார். அற்றம் என்னும் சொல் தடை எனப் பொருள்படும். அறிவு அனைத்துத் தடைகளையும் நீக்க வல்லது என்னும் உண்மையினை அறிவியல் உலகம் மெய்ப்பித்து வருகிறது. இன்று அறிவியல் வளர்ச்சியினால் விளைந்த தீமைகள் எனக் கருதப்படுபவை மனிதர்களின் கருணை இன்மையின் விளைவே ஆகும். அறிவியல் வளர்ச்சி இந்த அளவினை எட்டாதிருந்தால் ஊனமுற்றோரின் உலகம் ஒளியின்றி இருண்ட உலகமாகவே இருந்திருக்கும். இருண்டவீடு எனப் பாவேந்தர் எழுதிய நூலைப்போல் இருண்ட உலகம் எனும் ஒரு நூலினை எழுத வேண்டியிருந்திருக்கும். இன்று அந்த நிலை வளர்ந்த நாடுகளில் பெருமளவில் இல்லை. ஒரு பார்வையற்றவர் தகவல்களைத் தாமே இடுவரல் - விடுவரல் செய்யலாம், திருத்தலாம், நேர்த்தி யாக்கலாம், பிறருக்கு மின்னஞ்சலில் அனுப்பலாம். ஆங்கிலவழியில் ஓர் அந்தகர் இணையத்தில் நுழைந்து இதயத்து உணர்வுக்கு ஏற்றவாறு எந்தச் செய்தியைத் தேடினாலும், புதிது புதிதாக உதயமாகி அவர்தம் அறிவுக்கு விருந்தளிக்கிறது.

பிறர் எழுதிய நூல்களின் படமற்ற எழுத்துக்கள் அனைத்தையும் நொடிப்பொழுதில் படம்பிடித்துப் படித்துப்பேசும் தனிக்கருவிகளும், தனிச்சிறு கருவிகளும் கணினியுடன் இணைந்த மென்பொருள்களும் நடைமுறையில் கிடைக்கின்றன. சாலையில் வழிகேட்டு நடக்கவும், வாகனத்தில் செல்லும் போது வழிசொல்லவும் மென்பொருள் செயற்கைக் கோளுடனும், நிலப்படத்துடனும் இணைந்து செயல்படு கிறது. அண்மையில் ஒரு பார்வையற்றவர் வானொலி வழிகாட்டச் சீறிப்பாய்ந்து பந்தய உந்தினைச் செலுத்தியதைத் தொலைக்காட்சியில் பெரும்பாலோர் பார்த்திருப்பர். இவையெல்லாம் பல்வேறு கருவிகளின் மூலம் பெறத்தக்க அறிவுப்பார்வையாக உள்ளது. இந்த அறிவுப்பார்வை தமிழில் கிடைத்திடும் வழிகள் தவறும் நிலையிலேயே உள்ளன.

முதலில், ஆங்கிலவழியில் கிடைக்கும் கருவிகளை வாங்கிப் பயன்படுத்துவதில் தாங்கமுடியாத விலைத் துன்பம் உள்ளது. இதனை நீக்கிட விற்பனை செய்யும் நாடுகள், வாங்கும் நாடுகள் ஆகியவை வரிவிலக்கு அளித்தல், வளரும் நாடுகளில் அரசே அக்கருவிகளை வாங்கிச் சலுகை விலையில் வங்கிக் கடன், வள்ளல்களின் நிதியுதவி, கொள்வோரின் பங்களிப்பு ஆகியவற்றின் மூலம் விலைத்துன்பத்தைக் குறைத்து அந்தகருக்கு அவருக்கு உரிய கணினிக்கருவிகளைக் கொடுத்துவலாம். இப்போதே அவர் களுக்குரிய கணினிப் பயிற்சிகளை அளித்தால்தான் தனித்தமிழில் இக்கணித் தொழில்நுட்பத்தைத் தரும்போது அவர்கள் அயர்வின்றிப் பயன்படுத்துவர்.

மாற்றுப் பார்வைமுறை

காணாருக்கு மெய்யான ஒளிப்பார்வை உருவாக்குவதற்கு முன்பு மாற்றுப்பார்வை முறைகளைப் பின்பற்றித் தகவல் தொடர்பு வாய்ப்புகளை வழங்கலாம். மாற்றுப்பார்வை முறைகள் என்று இங்குக் குறிப்பிடப்பெறுவன செவித்திறன் மற்றும் தொடுவுணர்வு சார்ந்தவையாகும். முதலில் செவித்திறன் சார்ந்த ஏற்பாடுகள் குறித்தச் செய்திகளைக் காணலாம்.

செவித்திறனைக் கேள்விநிலை, கேளா ஒலிநிலை என இருநிலைகளில் பகுத்துக் காணலாம். கேள்வி என்னும் சொல்லை வள்ளுவர் அதிகாரப்பெயராக்கிச் செவிச்செல்வம் என்றே குறிப்பிட்டார். உண்மையில் பார்வையற்றோருக்குக் கல்வியும், செல்வமும் அதுவே ஆகும். முன்பு சொன்ன வெவ்வேறு தொழில்நுட்பங்களைப் பயன்படுத்திச் செவிவழித் தகவல் தொடர்பு உருவாக்கப்பட்டுள்ளது. இனிவரும் காலத்தில் மீச்சிறுவடிவில் மீத்திறன் கொண்ட செயற்கைக்கோளுடன் இணைந்து செல்பேசியில் செயல்படுமாறு தமிழில் தகவல் தொழில்நுட்பம் இடையறாது வழங்கப்பெற வேண்டும். கல்வி, தகவல் தொடர்பு, போக்குவரத்து, ஊடகப் பயன்பாடு ஆகியவற்றிற்கு உதவும் விதத்தில் இருவழித் தொடர்புள்ள வானொலிப் பணிகள் வழங்கப்பட வேண்டும்.

மந்திரக்கதைகளில் தேவதைகள் உடனிருந்து வழிகாட்டியதாகக் கேட்டுள்ளோம். ஆயின், இப்போது உள்ள தொழில்நுட்பத்தைப் பின்பற்றி எப்போதும் ஒரு துணைவர் உடன் இருந்து வழிகாட்டுவதைப் போல் உதவிடும் நிலையை உருவாக்கலாம். அப்படி உதவுவது ரோபோ என்றழைக்கப்படும் கருவி மனிதனாகவும் இருக்கலாம். இதற்கேற்றவாறு கருவிமனித உருவாக்க ஆய்வு, சப்பான் நாட்டினரின் கூட்டுத்தொழில்நுட்பத்துடன் விரைந்து இணைந்து தொடங்குதல் வேண்டும். சீனர் வன்பொருள் அறிவையும் மேல்நாட்டார் மென்பொருள் அறிவையும் அவர்தம் செல்வ வளத்தினையும் சேர்த்துப் பயன்படுத்தலாம். அவ்வாறு செய்வதன் மூலமே முறையான வழியில் முழுப்பயன் பெறலாம்.

ஒலிகளை இயல்பாகக் கேட்டிடும் நிலைகுறித்த ஆய்வுகள் ஒருபுறம் நடந்து வருகின்றன. மேலும், சிறப்பாக வாய் திறவாது பேசும்பேச்சினை ஒரு சிற்றொலிப் பெருக்கியை முன்கழுத்தில் பொருத்திச் செல்பேசியில் மறுமுனையில் தொடர்புகொண்டு வெற்றிபெற்றுள்ளனர். அமெரிக்கர், இதனை "துணைக்குரல் பேச்சு ஏற்புத் தொழில்நுட்பம்" (Subvocal speech, recognition) எனும் ஒரு துறையை உருவாக்கி ஆய்வு நிகழ்த்திவருகின்றனர். மேலும், கேளா ஒலி (Ultra Sound) எனும் பெயரில் இதுவும் தனித்ததோர் ஆய்வாக நிகழ்த்துவருகிறது. இயல்பாகச் செவிகள் கேட்கும் ஒலி அளவுக்கும் கீழே அதாவது 6 ஒலி அலகுக்கும் கீழ் உள்ள ஒலிகளைக் கேளாஒலி என்கின்றனர். இவற்றை ஈர்த்துக் கேட்டிடச்செய்யும் கருவியினை வடிவமைத்து வழங்குவதன் மூலம் அசையும் பொருள்களால் தொலைவிலிருந்து வரும் ஒலிகள், அசையாப் பொருள்கள்மீது எதிரொலித்துத் திரும்பும் ஒலிகள் ஆகிய இவற்றை, வெளவால் உணர்ந்து வாழ்வதுபோல் ஒவ்வொரு பார்வையற்றவரும் அவ்வொலிகளைக் கேட்டறியும் வாய்ப்பினை வழங்கலாம்.

"உருவம் கண்டுரைக்கும் தொழில்நுட்பம்" என்பதை ஆங்கிலத்தில் (இமேஜ் ரெகக்னிசன் அண்டு ஸ்பீச் இன்டெக்ரேடேசன்) என்று கூறுகின்றனர். இப்போது, எழுத்தினைப் பேச்சாக்கும் தொழில்நுட்பம் மட்டுமே உள்ளது. காணும் பொருள்கள் அனைத்தின் தன்மைகளைப் பேச்சுவடிவில் சொல்லும் மென் பொருள் உருவாக்கப்படும்பொழுது ஒலியினை உணர முடியவில்லை எனும் குறையைத் தவிர, தகவல் பரிமாற்றம் தடையின்றி நிகழும். இதனைச் செய்யவல்ல தொழில்நுட்பத்தைச் செவித்திறன் ஊக்கக் கருவியுடன் சேர்த்து வழங்கும்போது, ஒலிவழியில் ஒளிநிகர் தகவல் தொடர்பு உருவாகும். அதன் மூலமாகக் கண்ணொளியினால் பெறக்கூடிய தகவல்களைப் பெருமளவில் ஒலியினால் பெறலாம். இவ்விடத்தில் ரேடார் தொழில்நுட்பத்தைப் போர்ப் பயன்பாட்டிலிருந்து பொதுமனிதப் பயன் பாட்டிற்குத் தக்கவாறு மாற்றங்களைச் செய்தளித்தால் நுண்ணொலிகளைக் கேட்டுத் தகவல் அறியலாம்.

இப்போது நடைமுறையில் காது கேட்கும் கருவி உள்ளது. ஆயின், மேற்சொன்ன இவ்வாறான பல்வேறு தொழில் நுட்பங்களை ஒன்றிணைத்துப் புதியதொரு காதுகாக்கும்கருவி ஒன்றை உருவாக்கலாம். அக்கருவி சுற்றுப்புறத்தில் உள்ள பேரிரைச்சலைக் கட்டுப்படுத்தி வேண்டிய தகவல்களைத் தேர்ந்து பெறத்தக்கவாறு அமைத்திடல் வேண்டும். கணினி மயப்படுத்தப்பட்ட நுண்ணலை சார்ந்த காதுகேட்கும் கருவி உயர்நிலை ஆய்வுப் பயன்பாட்டில் அமெரிக்காவில் உள்ளது.

தொடுவுணர்வு சார்ந்த தகவல் பரிமாற்றத்தில் அதிர்வுகள், தொடுகுறியீடுகள் ஆகியவற்றை உருவாக்கும் கருவிகளால் தகவல்களை வழங்கலாம். நெற்றிக்கண் முறை(Forehead Retina) என்று ஜப்பான் நாட்டில் ஒரு தகவல் பரிமாற்றமுறை அறிமுகப்படுத்தப்பட்டுள்ளது. இது, புருவமத்தியில் நெற்றியின்மேல் மெல்லியதொரு உணர்வு முறையில் வரைவுகளை உருவாக்கித் தகவல்களைத் தருகின்றது. இத்துறையில் தொடர்ச்சியாக மேலாய்வுகளை நிகழ்த்திவருகின்றனர்.

கருவிப் பார்வை

கருவிப்பார்வை என்பது கண்ணாடி வடிவிலும் கண்ணுக்குள் பொருத்துகின்ற மீச்சிறு படக்கருவி வடிவிலும் பார்வை தருகின்ற ஏற்பாடுகளைப்பற்றிப் பேசுவதாகும். நியூயார்க் நகரில் ஜெரி என்பவருக்கு இவ்வாறு ஒருகண் பொருத்தப்பட்டுச் சோதனை செய்யப்பட்டது. அவர் கரும்பலகையில் உள்ள எழுத்துக்களைப்

படித்தார், தாமே பிறர் உதவியின்றி நகரில் சிறிதுதூரம் நடந்து சென்றார் எனும் செய்திகள் தொலைக்காட்சியில் காட்டப்பட்டு இணையத்திலும் இடம்பெற்றுள்ளன. நாளும் அறிவியல் என்னும் ஆங்கிலத்தில் அன்றாடம் வரும் மின் செய்தித்தாளில் இதுகுறித்த செய்திகள் இப்போது அடிக்கடி வருகின்றன.

"பயானிக்" எனப்படும் உயிர் மின்னணுத் தொழில்நுட்பத்தைப் பயன்படுத்திப் பார்வை வழங்கும் ஆய்வுகள் மேற்கொள்ளப்படுகின்றன. பிறவியில் ஏற்படும் பார்வையிழப்புக்குக் காரணமான மரபணு எதுவென அடையாளங் காணப்பட்டுள்ளது. அதனை மாற்றித் தீர்வுகாணும் முறை குறித்த ஆய்வு தொடர்கிறது. ஒளிஉணர் நரம்புகளை உயிர் நரம்புகளுக்கு மாற்றாகச் செயற்கை இழைகளை மீயுயர் தொழில்நுட்பம் கொண்டு உருவாக்கும் முயற்சியில் நாசா நிறுவனம் ஈடுபட்டு வருகின்றது. நுண்ணணு அளவை முறைகளைக் கொண்டு ஒளிஇழைகள் உருவாக்கப்படுவது இந்த ஆய்வின் மைய நோக்கமாகும்.

அமெரிக்காவின் கலிபோர்னியாவில் உள்ள "டோக்னி" மருத்துவமனை, "மாசாகூசிட்" தொழில்நுட்ப நிறுவனம் ஆகியவை இக்கருவிக்கண் உருவாக்கத்தில் இடையறாது ஈடுபட்டு வருகின்றன.

கலைஞர் கணினிக்கண் ஆய்வுமையம்

மேனாடுகளில் மேற்கொள்ளப்படும் கண்ணொளி வழங்கும் ஆய்வுத் திட்டங்களுடன் புதிய நோக்கில் அதுசார்ந்த ஆய்வுக் களங்களைக் கண்டறிந்து ஆய்வு நிகழ்த்தி மாற்றுத் திறனுடையோர் வாழ்வில் நல்லதோர் மாற்றம் செய்வோம் எனக்கூறும் கலைஞர் வாக்கினை நடைமுறைப்படுத்திட, பன்னாட்டுக் கலைஞர் கணினிக்கண் ஆய்வுமையம் எனும் ஒரு புதிய நிறுவனத்தை உருவாக்கவேண்டும். அந்நிறுவனம் படிப்படியாகக் கருவிக்கண் மற்றும் மெய்யொளிப் பார்வையை வழங்குதல் குறித்த ஆய்வுகளை உயர்தொழில்நுட்ப அறிவுடைய மருத்துவர்கள், பொறிஞர்கள், ஆகியோரைக்கொண்டு நிகழ்த்த வேண்டும். இதுவன்றிப் பிற வகையில் திறன்இழந்த கேளார், கால்முடம்பட்டோர், மனவளங் குன்றியோர் ஆகியோரின் துயர்துடைக்கவும் அந்நிறுவனத்தார் பணியாற்ற இயலும். மரபணுத் தொழில் நுட்பம் ஆகிய ஆய்வுகளைக் கொண்டு கல்லீரல், இருதயம் முதலிய செயற்கை உறுப்பாக்க ஆய்வுகளை மேற்கொள்வதன் மூலம் பொதுமனிதச் சமுதாயத்திற்கே தொண்டாற்றலாம்.

நல்லோர் பலரும் நோயின்றி வாழ வழி செய்யலாம். இத்தகைய சிந்தனைகள் வரும் தேர்தல் அறிக்கையில் இடம்பெறும் அளவிற்கு இக்கருத்திற்குச் சிறப்பிடம் அளிக்க வேண்டும்.

Role of Cloud Computing in Tamil Language Development

Mrs. R. RajaRajeswari

Assistant Professor in Computer Science
M.V.M Government Arts College for Women,
Dindigul
ganesh_akash@yahoo.com

Dr. Mrs. A. Pethalakshmi

Head & Associate Professor in Computer Science
M.V.M Government Arts College for Women,
Dindigul
alaguvisu@gmail.com

Cloud Computing is a new computing paradigm which is expected to transform the way computing is done today, in near future. Cloud Computing offers virtualization of all high end Computing Services. And it offers four layers of Services:

SaaS (Software as a Service), PaaS (Platform as a Service), IaaS (Infrastructure as a Service) and SaaS (Software as a Service). These services can cut the software cost, storage cost and utility cost of running a wide computer network. This research paper addresses how cloud computing can facilitate Tamil Language development. The following areas are identified as prospective avenues of development: E-learning, Tamil Computing, Tamil E-resources and Tamil Internet Services. This research paper will analyse the existing tamil software services in the above said areas and explore these avenues of tamil development through Cloud Computing.

(Keywords: Cloud computing, E-Learning, Tamil Computing, Tamil E-Resources)

1. Introduction

Forecast on Computing tells cloudy days are ahead for Computer Scientists. Software Developers and Users should be ready for a changeover as Cloud Computing waves roll down into the Internet. This new computing paradigm is expected to transform the style of computing in a remarkable way. Tamil, one of the world's five classical languages, is a privileged language, to be learnt by Lord Buddha[11]. Its history dates back to its existence on stone inscriptions in Jerusalem. This research paper tries to make these two ends meet ie. how the new computing paradigm, cloud computing, can be connected with this historical language. This research paper is organized as follows : an introduction to cloud computing, avenues of Tamil Language Development facilitated by cloud computing and the final conclusion.

2. An insight into Cloud Computing

Cloud Computing [8] is Internet based computing, whereby shared resources, software and information are provided to computers and other devices on demand like a public utility.

Cloud computing [9] is Internet based ("Cloud") development and use of computer technology ("computing"). The cloud is a metaphor for the Internet based on how it is depicted in computer

network diagrams and is an abstraction for the complex infrastructure it conceals. According to a 2008 paper published by IEEE Internet Computing “Cloud Computing is a paradigm in which information is permanently stored in servers on the Internet and cached temporarily on clients that include desktops, entertainment centers, table computers, note books, wall computers, hand holds, sensors, monitors etc.”

Cloud computing offers virtualization of all high end services and it offers four layers of Services[1,2,4,5,6,7].



Courtesy : www.cloudtweaks.com

Fig.1

(i) Applications or Software as a Service

SaaS means delivering software over the internet . This nullifies the need to install and run application on individual computers, making software upgrades, maintenance and support obsolete. (e.g) Salesforce.com

(ii)Platform as a Service

PaaS is the delivery of a computing platform as a service . PaaS offerings facilitate deployment of application without the cost and complexity of buying and managing the underlying hardware and software and provisioning hosting capabilities. (e.g) Windows Azure, Amazon Elastic Computer Cloud.

(iii)Infrastructure as a Service

IaaS is a provision model is which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components. The service provider owns the equipment and is responsible for housing, running and maintaining it. The client typically pays on a per-use basis. (e.g) Amazon S3

(iv)Computing as a Service

CaaS [1] integrates the above said services. (e.g) Verizon’s CaaS .Like other cloud offerings, Verizon’s CaaS allows customers pay for data-center resources such as storage and application hosting dynamically based on the amount of resources they consume. These services can cut the software cost, storage cost and utility cost of running a wide network. The following sections bridges Cloud Computing Services and Tamil Language Development in prospective avenues.

3. Tamil E-resources

Literary Resources in a language symbolises its richness and liveliness. Now in this internet age, e-mail box has replaced the conservative letter box in a house. Hence E-resources have a definite role in making a language flourish. Internet has paved a way for Tamil Internet magazines, e-books and also a Tamil electronic library (www.tamilelibrary.org). For example, www.projectmadurai.com plays a major role in the area of e-documentation of old Tamil literature. One has free access to Tamil Literature in this website from “Abirami Anthathi” to “Alai Osai”. Apart from this, many tamil magazines have their own websites comprising even digital archives of old ones. (e.g) www.vikatan.com. Also any creator’s work needs to be published and recognised, for both the growth of the language and the creator.

Tamil Internet magazines act as writer’s workshops, for budding Tamil writers from which they can learn and also where they can publish their work. (e.g.) www.thinnai.com, www.nilacharal.com. With the advent of Cloud Computing, usage of IaaS can cut storage costs, emancipating separate clouds for old Tamil Literature, Magazines and modern Tamil Literature. Such e-resources available on the Internet facilitates free access to Tamil Literature for any one across the globe, which is a requisite for a language to grow.

4. Tamil E- Learning

Tamil community has settled across the globe even many decades before. And there is need to teach Tamil for the next generation young learners to keep the tamil heritage and language expand its horizon. Internet takes the avatar of a virtual Guru for this purpose. “International Academy for Internet Tamil “

Formerly, Tamil Virtual University (www.tamilvu.org) does this job with ease in collaboration with Tamil University, Thanjavur and awards Degrees, Certificates and Diplomas in Tamil for Tamil learners across the globe through Internet.

Virtual Class room, a part of this academy’s website, which transits the learner from the living room into a class room needs additional software in a Remote terminal, Eg. Multimedia Software. Using PaaS, this software can be downloaded and any remote user can turn the terminal into a virtual class room. And apart from this enlightening Video lectures by eminent Tamil Professors will reach across the globe when available as AaaS for enthusiastic learners.

5. Tamil Internet Services

Twenty first century is the century of Information and Communication Technology. At the click of a mouse, Internet provides both communication in its highest speed and information in its best shade ,at any moment. Searching for Tamil web pages is mostly done in English and only some Tamil search Engines like www.googletamil.com. does this job in Tamil that too using transliteration method. Tamil e-mail software like www.azhahi.com though exist are not popularly used by tamil population. Since personal and official communication is under the process of transformation from letters to mails, sending mails in mother tongue will be a necessity among users who have a fast pace of life, to share their true emotions.

Apart from this, a language should be put into use in all forms of communication by that language speaking people for its longevity and the one not, can exist only in books and stone inscriptions. And CaaS comes as a help at this moment for the software developers to develop next generation Tamil search

Engines and Tamil E-mail sending software. Required Infrastructure, Platform, and Computing for developing these software can be got from IaaS, PaaS and CaaS providers respectively.

6. Tamil Computing

Tamil Computing provides Tamil software tools which are used in a day to day basis. (e.g) Tamil word processors, Tamil Database software like Amudham These tools need to get popularised and put into use by software users. Also Tamil Speech Recognition software is yet to come. To develop Tamil Speech Recognition Software more computing power is needed. To develop or run Speech Recognition applications, because of digital filtering and signal processing high processing speed and more memory are required[3]. And cloud computing comes handy by providing them through CaaS, since cloud computing transforms a desktop into a super computer. To help website developers develop tamil websites easier, Tamil website design templates can be made available as AaaS .

7. Conclusion

Tamil is a unique language, an universal language, whose literature suits mankind of all centuries, all places and all tribes. [10] states that context sensitive rules of modern computer science are found in Tholkaapiyam, a classical literature. This language has undergone transformations, its storage from stone inscriptions to palm leaves, paper and now Web pages.

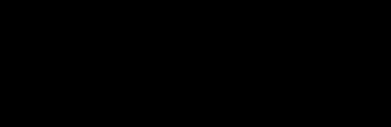
For a language to live long, it should be spoken without the acquaintance of any foreign language. Today's tamil language is spoken along with many english words. In spite of this, this language is ever young and vibrant and can be symbolised as a wild flower which blossoms in a natural way. Hence this language will evolve with any emerging technology and tamil language can develop in the above said avenues through cloud computing services .

References

1. <http://connectedplanetonline.com>
2. <http://edgewater.tech.wordpress.com>
3. <http://www.faqs.org/docs/Linux/SpeechRecognition.Howto.html>
4. <http://www.scribd.com>
5. <http://searchcloudcomputing.techtarget.com>
6. <http://universitybusiness.com>
7. <http://web2.sys.con.com>
8. en.wikipedia.org
9. Dr. Durgesh Pant et.al, Cloud Computing, CSI Communication, January 2009.
10. Dr. Gift Siromeny, Context Sensitive Rules in Tolkappiam, Proceedings of the Second World Tamil Conference, 1968
11. News, Makkal TV, Jan 2009.

14

தமிழில் தேடு பொறிகள்



தமிழ் இணையத்தளமும் தேடு பொறிகளும்

பெரி. கபிலன்

உறுப்பினர் மன்றம், உத்தமம், உதவிப் பேராசிரியர், கணினி அறிவியல் துறை,
மதுரை காமராசர் பல்கலைக்கழக கல்லூரி, மதுரை-625002. spkabi@gmail.com 09894406111

இரா.கார்த்திக்

ஐ.பி.எம்., இந்தியா, ஓயிட் பீட்டு, பெங்களூர். karthick.krishnan@in.ibm.com 09731809067

கருத்து அறிமுகம்

தமிழ் மொழியின் செம்மை அதன் தொன்மையில் மட்டுமல்லாது அதன் தொடர்ச்சியிலும் உள்ளது. தொன்மைத் தமிழின் தொடர்ச்சிக்கான சாத்தியக் கூறுகளை, இணையத்தோடு தமிழ்மொழி இணைந்து செயலாற்றுவதன் மூலமே உருவாக்க இயலும். இணையம் என்கிற இயங்குதளம் வழங்கிவருகிற வாய்ப்புகளை சரியாகப் பயன்படுத்திக் கொள்ளவும், தமிழில் உள்ள இணையத்தளங்களை தற்கால தேடு பொறிகளின் தன்மைகளுக்கேற்ப தகவமைப்பதற்கான அவசியத்தையும், ஆலோசனைகளையும் இக்கட்டுரை முன்வைக்கிறது.

கண்டுபிடிப்புகளின் வெற்றியோ, வீழ்ச்சியோ அது அனைத்துத் தரப்பு மக்களிடம் ஏற்படுத்தும் பயன்களைப் பொறுத்தே அமைகிறது. இந்த வகையில் இணையம் என்கிற கண்டுபிடிப்பு பெற்ற வெற்றியை அதன் பயன்பாடுகளைக் கொண்டே அளவிடலாம். இணையம் பயன்பாட்டிற்கு வந்த பிறகுதான் தகவல் தொழில்நுட்பம் என்ற துறையே உலகிற்கு அறிமுகப்படுத்தப்பட்டது. இந்தத் தொழில்நுட்பம் உலகெங்கிலும் உள்ள கணினிகளை இணைத்து அறிவுப்புரட்சிக்குக் காரணமானது. அதே சமயத்தில் உரிய இணையத்தளங்களைத் தேடுவதில் சிக்கல்கள் உருவாகின. இந்தச் சூழலில் தான் தேடு பொறிகள் அறிமுகமாயின. இணையத்தின் இப்பெரும் வளர்ச்சிக்கும் தேடு பொறிகளே தூண்களாகின.

இந்தத் தேடு பொறிகள் செயலாற்றும் விதம் மனித மூளையின் செயல்பாட்டிற்கு ஒப்பானதாகும். ஒருவரைப் பற்றி நினைத்த உடனே அவர் தொடர்பான செய்திகள் நினைவில் விரிவது போல் இந்தத் தேடு பொறிகளில் தேடலுக்குத் தேவையான உள்ளீடுகள் அளிக்கப்பட்ட அடுத்த கணமே, உள்ளீடுகளுக்கேற்ப தகவல் அடங்கிய இணையத்தளங்களை பகுத்துப் பட்டியலிட்டு விடுகின்றது. இந்த மின்னல்வேகச் செயல்பாடு தான் அனைவரும் இணையத்தை விருப்பத்துடனும், நம்பிக்கையுடனும் பயன்படுத்த அடிப்படைக் காரணங்கள்.

பன்னெடுங்காலமாய் வழக்கில் உள்ள மொழியை இணையத்தோடு இணைப்பதிலும், அதன் பயன்பாட்டைப் பரவலாக்குவதிலும் தேடு பொறிகளின் பங்கு அளப்பறியது. தேடு பொறிகளின் தன்மைகளையும், அவை செயல்படும்விதத்தையும், அவற்றை உருவாக்கும் விதிகளையும் ஆராய்ந்து அதற்கேற்ற வகையில் இணையத்தளங்களில் தகவல்களையும், திறவுச் சொற்களையும் உள்ளீடு செய்வதில் தான் நம் மொழியின் நீட்சி உறுதி செய்யப்படுகிறது.

தேடு பொறிகள் தமிழில் தேடுவதற்கான சாத்தியக் கூறுகளை உருவாக்கியுள்ள போதிலும் தமிழிலேயே தேடுவதென்பது இன்றும் முழுமை பெறாத ஒன்றாகவே இருந்து வருகிறது. தமிழில் இணையத் தளங்களை உருவாக்கும் பொழுதே பின்பற்ற வேண்டிய விதிமுறைகளையும், தேடு பொறிகள் சரியாகத் தொகுக்கும் வகையில் தளங்களை உருவாக்கும் விதிமுறைகளையும் பற்றியதே இந்த ஆய்வுக் கட்டுரை.

தகவல் தொழில்நுட்ப புரட்சியும், இணையமும் உலக மொழிகளின் மேல் தொடர்ந்து அதிர்வுகளை செலுத்தி வரும் ஒரு சூழலில், இந்த செம்மொழி மாநாடு, மொழி மிதான நம் பார்வைகளுக்கு மேலும் தெளிவு சேர்க்கும் வகையில் அமைந்துள்ளது. மொழியின் தொடர்ச்சிக்கான அணுகுமுறைகளை தொழில்நுட்பப்பிரிதலுடனும் தர்க்கரீதியாகவும், ஆய்வு செய்யும் களமாகவும் இது போன்ற நிகழ்வுகள் அமைய வேண்டும். இணையத்தில் தமிழ் மொழியின் பரப்பையும், பயன்பாடுகளையும் அதிகரிக்க, தொழில்நுட்பப்பிரிதலுடன் கூடிய அணுகு முறைகள் மொழியாளர்களுக்கு அவசியம் என்ற கருத்தையும் இந்தக் கட்டுரை முன் வைக்கிறது.

இணையத்தில் தகவல் தேடுவோரின் எண்ணிக்கை முன் எப்போதும் இல்லாத அளவில் உயர்ந்து வருவதையும், அதே சமயம் தமிழில் தகவல்தளங்கள் அதிக அளவில் பழக்கத்தில் இருந்தாலும் தமிழிலேயே தகவல்களைத் தேடி சரியான தகவல்களைத் தருவிப்பதில் அதிக அளவில் நடைமுறைச் சிக்கல்கள் இருப்பதும் நம் அறிந்ததே. இவ்வாறான சிக்கல்களை தொழில்நுட்பத்தின் மூலம் சரிசெய்து, நமது மொழியின் தொடர்ச்சிக்கான சாத்தியக்கூறுகளை ஆராயவேண்டும். மேலும், இது தொடர்பான விழிப்புணர்வை அறிவியாளர்கள் மட்டுமல்லாது மொழியாளர்கள் மத்தியிலும் பதிவு செய்ய வேண்டிய நேரமிது.

இணையம் போன்ற ஒரு இயங்குதளத்தில் ஒரு மொழி இயங்கும் தன்மைகளை ஆராய்ந்தால் தான் மொழியின் ஆயுளை நீட்டிக்க நாம் செயல்படுத்தவேண்டிய செயல்களும் புலனாகும். பயன்பாடுகளைப் பொறுத்தே மொழியின் வளர்ச்சியோ வீழ்ச்சியோ அமையும் என்பது பொதுவான உண்மை. இதனை மனதில் கொண்டு இணையத்தில் உள்ள தகவல் தளங்களையும், தகவல் தேடுவோரின் செயல்பாடுகளையும் உற்று நோக்கினால் இன்னுமொரு உலகலாவிய உண்மை விளங்கும். தகவலைப்பெற இணையத்தை நாளும் யாவரும் தொடங்கும் இடம் தேடு பொறிகள் தான்.

தகவல்தேடுவோர் உள்ளீடு செய்யும் வார்த்தைகளுக்கு ஏற்பத் தகவல்தளங்களை வரிசைப் படுத்தி, இணையத்தில் தகவல் தேடலை எளிதாக்கி இன்றளவில் தகவல் தேடுவோருக்கான நுழைவாயிலாக இந்த தேடு பொறிகள் விளங்குகின்றன. மேலும் இணையத்தள முகவரியே இல்லாமல் தகவல் தேட துவங்குவோர்க்கு இந்த தேடு பொறிகள் தான் நன்றம்பிக்கை முனைகள்.

தேடு பொறிகள் அமைப்பு, செயல்முறை ஆகியவற்றைத் தெரிந்து கொண்டு அதற்கேற்ப தமிழிலேயே தகவல் தளங்களை உருவாக்குவதும், தகவமைப்பதுமே தமிழ் மொழியின் பரப்பை இணையத்தில் அதிகரிப்பதற்கு நாம் செய்யும் முதன்மையான பணியாக இருக்க முடியும். அதன் பொருட்டு நமக்கு தேடு பொறிகளை பற்றிய அறிமுகமும், தகவலை தொகுத்து மற்றும் பகுத்து வழங்கும் முறை பற்றியும் அறிவது அவசியமாகிறது. இணையம் உருவாக்கியுள்ள வலைப்பின்னலும் கணினிகளின் கூட்டமைப்புமே தேடு பொறிகள் செயல்படுவதற்கான சாத்தியத்ததை உருவாக்குகின்றன. தேடு பொறிகள் இந்த வலைப்பின்னலில் அமைந்துள்ள எல்லா இணையத்தளங்களையும் அறிந்து கொண்டிருக்கும் வகையில் அமைக்கப் பெற்றுள்ளன.

தேடு பொறிகளின் செயல்பாடுகளை மணித மனத்தின் செயல்பாடுகளோடு ஒப்பிடுவதன் மூலமே விளக்கிவிடலாம். இது ஒரு மணிதரைப் பற்றி நாம் நினைத்த உடனே அவர் தொடர்பான எண்ணங்கள் நம் மனத்திரையில் விரிவதற்கு ஒப்பாகும். இந்த தேடு பொறிகளின் தகவல் தொகுக்கும் வேகமும் அவை நமக்குத் தகவலை அளிக்கும் விதமும் தான் தேடு பொறிகளுக்குப் பெரும் வரவேற்பை இணையத்தில் உருவாக்கியுள்ளன. இந்தத் தேடு பொறிகளின் வேகத்திற்கு அவற்றின் பின்புலத்தில் அமைந்துள்ள கணினிகளின் ஆற்றல் தான் காரணம் என்றாலும், நமக்கு ஆச்சரியமளிப்பவை தேடுபொறிகளுக்கு மூளையாகச் செயல்படும் "வெப் கிராவ்லர்களின் (Web Crawler)" செயல்பாடே. தேடு பொறிகள் தகவல் தளங்களுக்குத் தங்கள் பிரதிநிதியாக இந்த "வெப் கிராவ்லர்ஸ்" அனுப்புகின்றன. வெப் கிராவ்லர்ஸ் அந்த குறிப்பிட்ட தளம் பற்றிய தகவல்களையும், திறவுச் சொற்களையும் தொகுத்து தேடு பொறிகளின் "தரவுத்

தளங்களில் (Data Base)" பதிவுசெய்து தொடர்ந்து அந்த தளங்களில் எந்த மாறுபாடுகள் நிகழ்ந்தாலும் கண்காணித்து அவற்றையும் பதிவு செய்து கொள்கின்றன. தகவல்தளங்களையும் தேடு பொறிகளையும் இணைக்கும் புள்ளிகளாக அந்த குறிப்பிட்ட தளங்களில் இருந்து "வெப் கிராவ்லர்ஸ்" மூலம் பதிவு செய்யப்பட்ட திறவுச் சொற்களே அமைகின்றன. திறவுச் சொற்கள் தான் தகவல் தளங்களை தகவமைப்பதில் முக்கிய பங்கு வகிக்கின்றன. இந்த திறவுச் சொற்களை தெரிவு செய்து தருவதில் மொழியாளர்களின் பங்கு அவசியமாகிறது.

திறவு சொற்கள் கொண்டு தகவமைத்தல்

திறவு சொற்களை தேர்ந்தெடுத்தல்: ஒரு புதிய இணையத்தளத்தை நிறுவும் பொழுது, மொழியாளர்கள் மற்றும் தகவல் உள்ளீடு செய்வோர் இணைந்து, அந்த இணையத்தளத்தோடு தொடர்புடைய வார்த்தைகளையும், பயனீட்டாளர்கள் தேடலுக்கு பயன்படுத்தக்கூடிய வார்த்தைகளையும், தேர்வு செய்து அவற்றை திறவு சொற்களாக பயன்படுத்த வேண்டும்.

தேர்வு செய்த திறவு சொற்களை சரியாக பயன்படுத்துதல்: தேர்வு செய்யப்பட்ட திறவு சொற்களை தலைப்புகள், தள முகவரி, கோப்புகளின் பெயர்கள், படங்களின் பெயர்கள், பத்தி தலைப்புகள் ஆகியவற்றில் பயன்படுத்த வேண்டும். அதே சமயத்தில் இந்த திறவு சொற்கள் ஒரே இடத்தில் தொடர்ந்து வராதவாரும் அமைவது நலம்.

திறவு சொற்களின் அடர்த்தி: திறவு சொல் அடர்வு என்பது வலைத்தளத்தின் ஒரு பக்கத்தில் திறவுச் சொல் எத்தனை முறை இடம் பெற்றுள்ளது என்பதற்கும் அந்த பக்கத்தில் அமைந்துள்ள மொத்த வார்த்தைகளின் எண்ணிக்கைக்கும் உள்ள ஒப்பீடே ஆகும். உதாரணமாக 1000 வார்த்தைகள் கொண்ட தகவல் பக்கம் ஒன்றில் 100 முறை திறவு சொல் அமையப் பெற்றிருந்தால் திறவுச் சொல் அடர்வு 10 சதவீதமாகும். இந்த திறவு சொற்கள் இணையத்தளத்திற்கு வரும் பயனிட்டாளர்களின் திரையில் தெரியாது. இந்த திறவு சொற்களை பயனீட்டாளர்கள் அறியாத வண்ணம் பயன்படுத்துவதே இதன் நோக்கமாகும். இது எந்த வகையிலும் இணையத்தளங்களின் தரத்தை பாதிக்காது.

தேடுபொறிகளுக்கு ஏற்ப தகவமைப்பதற்கான படி நிலைகள்

- தகவல் தளங்களின் முதல் பக்கங்களின் மீதே தேடு பொறிகளின் கவனம் 60 சதவீதத்திற்குமே மேலாக இருக்கும். எனவே இந்த பக்கத்தை அதிக கவனத்துடன் கையாள வேண்டும்.
- திறவு சொற்கள் மூலம் இணையத்தளங்களைத் தகவமைக்க வேண்டும்.
- இணையத்தளங்களில் உள்ள திறவு சொற்களை தேடுபொறிகள் எளிமையாக பின்பற்றுமாறு அமைக்க வேண்டும்.
- ஒரே நோக்கத்தோடு நிறுவப்பட்ட தளங்களோடு தொடர்பு கெள்ளுமாறு இணையத்தளங்களை அமைக்க வேண்டும்.
- நமது தளத்தின் பெயர் பிற தளங்களில் வரிசை படுத்தப்பட ஆவண செய்தல் வேண்டும்.
- இணையத்தளங்களில் தொடர்ந்து உள்ளீடு செய்தல் வேண்டும்.
- தேடு பொறிகளின் பிரதிநிதிகளை எல்லா இடங்களிலும் எளிதாக உலவும் வகையில் வலைப் பக்கங்களை அமைக்க வேண்டும்.
- இணைய வழித்தடங்களை (SITE MAP) இணையத்தளத்தில் சேர்க்க வேண்டும்.
- தொடர்ந்து இணையத்தள செயல்பாடுகளைக் கவனித்து அதற்கேற்பத் தளத்தின் முக்கிய வார்த்தைகளையும் பயனீட்டாளர்களுக்குத் தேடும் பக்கங்களை செம்மைப்படுத்த வேண்டும்.
- தேடு பொறிகளை ஏமாற்ற முயற்சிசெய்யக்கூடாது

இணையத்தில் பிற மொழி தகவல் தளங்களும் பல்கிப்பெருகி வருவதனால் தமிழ் மொழியில் அமைந்துள்ள தளங்களை ஆங்கிலத்தில் தேடும் ஒரு சூழலும் உருவாகிறது. தேடு பொறிகளுக்கு ஏற்ப இணையத்தளங்களைத் தகவமைக்கும் நுட்பம் வியாபாரரீதியாக பெரும் வெற்றி அடைந்ததுள்ளது. முன்னனி நிறுவனங்கள் வணிகரீதியாக பலன்களைக் கருத்தில் கொண்டு இந்த நுட்பத்தைத் தொடர்ந்து பயன்படுத்தி வருகின்றன.

மொழியை முன் நிறுத்தும் முனைப்போடு இந்தத் தொழில்நுட்பத்தை அணுகும் ஆய்வுகள் மிகச் சொற்பமாகவே நிகழ்ந்து வருகின்றன. முன்னனி தேடு பொறிகளில் மொழியின் கூறுகள் ஆராயப்பட்டு அவற்றிற்கேற்ப மொழிக் கருவிகள் உருவாக்கப்பட்டு வருகின்றன. இந்தக் கட்டுரை மொழியின் பல்வேறு கூறுகளை கருத்தில் கொண்டு தகவல் தொழில்நுட்பப் பரப்பில் தமிழ் மொழியை எல்லா நிலைகளிலும் இணைப்பதற்கான சாத்தியக்கூறுகளை ஆராய்ந்து அவற்றில் முக்கியமானதும், எளிமையானதுமான இணையத்தளங்களில் திறவுச் சொற்களின் மூலம் தகவமைத்தல் பற்றிய கருத்துக்களை முன் வைக்கிறது. தேடு பொறிகளின் பட்டியலில் தரமான தமிழ் இணையத்தளங்களை திறவுச் சொற்களை கொண்டு தகவமைத்தல் மூலம் முதன்மை பெற செய்வதே மேற் படி ஆய்வின் நோக்கமாகும்.

இடைமுகத் தமிழ் உள்ளீட்டு மென்பொருட்கள்

—ஓர் ஒப்பீடு

Pannirukaivadivelan R

University of Madras (pann1973@gmail.com)

முன்னுரை

மைக்ரோசாப்ட் வேர்ட் ,பேஜ் மேக்கர் ,போட்டோஷாப் ,இன்டிசைன் ,க்யூயார்க் எக்ஸ்பிரஸ், கோரல் டிரா, ப்ளாஷ் போன்ற மென்பொருட்கள் ஆங்கிலத்தைக் கணினியில் நேரடியாக உள்ளீடு செய்ய உதவுகின்றன. இம்மென்பொருட்கள்மூலம் பிற மொழிகளைக் கணினியில் உள்ளீடு செய்ய உரிய இடைமுக மென்பொருள் தேவைப்படுகிறது. இவ்வகையில் தமிழ் உள்ளீட்டுக்கு உதவுவனவான L I P, முரசு அஞ்சல், இ-கலப்பை, என்.எச்.எம்., புதுவை தமிழ் எழுதி, துணைவன், அரும்பு, குறள் ,வானவில் ,தமிழ் விசை போன்ற இடைமுகப்புகள் அமைகின்றன .இத்தகு மென்பொருள்கள் இடைமுகத் தமிழ் உள்ளீட்டு மென்பொருள்கள் எனப்படுகின்றன. இம் மென்பொருள்களை அறிமுகப்படுத்தி, இவற்றின் செயல்பாட்டை விளக்கும் நோக்கில் இக்கட்டுரை அமைகிறது.

L I P

மைக்ரோசாப்ட் குழுமம் லிதுவானியன்)Lithuanian(, செர்பியன்)Serbian(, இந்தி)Hindi(, மராத்தி)Marathi(, தமிழ்)Tamil,(தாய்)Thai (போன்ற மொழிகளைக் கணினியில் கையாள Language Interface Pack (LIP)ஐ வழங்கியுள்ளது. இதை மைக்ரோசாப்ட் நிறுவனத்திடம் அனுமதி பெற்றுப் பயன்படுத்தலாம்.

அழகி

இது தமிழ் மென்பொருள்களில் தனித்தன்மை வாய்ந்த ஒன்று ஆகும். இதை விண்டோசின் அனைத்துச் செயலிகளிலும் நேரடியாகவே தமிழை உள்ளீடு செய்யப் பயன்படுத்த முடியும். இதில் யுனிகோட் (Unicode), திஸ்கி ஆகிய எழுத்துருக்களை ஒலியியல், தமிழ்நெட் 99, தமிழ் தட்டச்சு போன்ற விசைப்பலகைகள்மூலம் தட்டச்சு செய்யமுடியும். உலகின் முதல் 'இருதிரை' ஒலிபெயர்ப்புக் கருவி கொண்ட இது, ஒரு எழுத்துருவிலிருந்து இன்னொரு எழுத்துருவிற்கு மாற்றும் வசதி படைத்தது.) MS-Word, Excel, powerpoint, Access, Pagemaker, Photoshop, Outlook Express, MSN messenger(போன்ற எல்லாச் செயலிகளிலும் இதை நேரடி ஒலிபெயர்ப்புவழி உள்ளீடு செய்யலாம். மேலும், அனைத்து திஸ்கி (TSCii), டாப்)Tab) மற்றும் யுனிகோட் எழுத்துருக்களுக்கு மாற்றம் செய்யலாம்.

இப்படி மாற்றும்போது ஒரே நேரத்தில் பல நூறு திஸ்கி கோப்புகளை மாற்றும் வசதி படைத்தது இது. அத்தகைய வசதிக்குப் பெயர் bulk Unicode converter ஆகும் .இது விரிவான யுனிகோடு உதவிக் கோப்புகளை உள்ளடக்கியது.

விசைத்தமிழ்

உலகில் முதல் தமிழ்ச் சொற்பிறைச்சுட்டியைக் கொண்டது, இம்மென்பொருள். கணினியில் தமிழின் பயன்பாட்டை அனைவரும் உணர்ந்து செயல்படவைக்கும் முதன்மையான மென்பொருள் இதுவாகும் . இம்மென்பொருள் 23இலட்சம் தமிழ்ச்சொல் தொகுப்பைக் கொண்டுள்ளது.

இது எல்லா வகையான எழுத்துருக்களிலும் சொற்பிழைகளைக் கண்டறியும் சிறப்புப் பெற்றது .இதில் எல்லா வகையான விசைப்பலகைகளையும் பயன்படுத்தமுடியும். உள்ளீடு செய்யும்போதே தவறான சொற்களையும் ,இலக்கணப்பிழையினையும் காண்பிக்கும் வசதி இதில் உண்டு .இது காண்பிக்கும் சரியான சொற்களைக் கொண்டு தவறான சொற்களை நீக்கி விடலாம் .ஒரு நொடிக்கு 80முதல் 120சொற்களின் சொற்பிழையினைக் கண்டறியும் திறனுடையது.

விசைத்தமிழ் Microsoft office word, Microsoft office excel, Microsoft office power point, Microsoft office front page, Adobe pagemaker, Adobe Photoshop போன்ற அனைத்து மென்பொருளிலும் தமிழில் உள்ளீடு செய்ய உதவுகிறது.

1999 ஆம் ஆண்டில் உருவாக்கப்பட்ட யூனிகோடை, 2000 மற்றும் அதற்குப் பின்னர் வடிவமைக்கப்பட்ட இயங்குதளத்தில் (operating system) மட்டுமே பயன்படுத்த இயலும் .அதாவது windows 98, windows போன்றவற்றில் அதைப் பயன்படுத்த இயலாது. இக்குறைபாட்டினை விசைத்தமிழ் நீக்குகிறது. மேலும், யூனிகோட் முறை பயன்படுத்தமுடியாத பழைய இயங்கு தளங்களில் யூனிகோட் முறையிலுள்ள கோப்புகளை அவரவர்களிடமுள்ள எழுத்துருக்களில் பார்வையிடுவதற்கேற்ற மென்பொருள் கருவி இதில் இணைக்கப்பட்டுள்ளது. இது பிற மொழியிலுள்ள தகவல்களை மாற்றும் தன்மையுடையது . எழுத்துப்பெயர்ப்பு அல்லது வரிவடிவ மாற்ற Transliteration or script conversion என்னும் வசதியும் இதில் உண்டு.

தமிழ்ச் சொற்பிழைசுட்டி மென்பொருள்

ஆங்கிலத்தில் தரமான சொற்பிழைசுட்டிகள் 1990களில் அறிமுகம் ஆயின .ஆங்கில மொழிக்கான இம்மென்பொருட்கள் கணினியில் ஆங்கிலத்தை உள்ளீடு செய்யும்போது பிழையின்றி உள்ளீடு செய்ய உதவுகின்றன . இதுபோல் தொல்காப்பியம் ,நன்னூல் இலக்கணங்களை அடிப்படையாகக் கொண்டு தமிழ் சொற்பிழைசுட்டி மென்பொருள் வடிவமைக்கப்பட்டுள்ளது .இது தனித்து இயங்குவதோடு Notepad, wordpad ஆகியவற்றிலும் இயங்கும் தன்மை கொண்டது.

இ-கலப்பை 2.0 (e-kalappi – 2.0)

ஒரு கணினியில் உள்ளீடு செய்த தமிழ் உரையை, மற்றொரு கணினியில் ஏற்றுக்கொண்டு, எழுத்துரு காரணமாகப் படிக்க இயலாது. இதனைப் போக்க Unicode தேவைப்படுவதுபோல், இ கலப்பையை முகுந்தராசு என்பவர் உருவாக்கினார். யூனிகோடு வடிவில் தட்டச்சு செய்யும் விசைப்பலகையே இ.கலப்பை ஆகும். இது 2001-2004 ஆம் ஆண்டுகளில் தமிழ்த் தட்டச்சினை எளிமைப்படுத்தியது. இதுவே பின்னர் என்.எச்.எம். என்னும் மென்பொருளுக்கு முன்னோடி எனலாம். இது விண்டோஸ் செயலிகளில் TSC எழுத்துருவிலும் செயல்படுகிறது.

என்.எச்.எம்)NHM)

என்.எச்.எம். எழுதி, என்.எச்.எம் எழுத்துருமாற்றி என இரு வகை மென்பொருள்கள் உள்ளன. NHM என்பதன் விரிவாக்கம் New Horizon Media ஆகும். இது 2004 ஆம் ஆண்டில் பத்ரி சேசாத்திரி, நாகராஜன், சத்யநாராயணன், ஆனந்தகுமார் ஆகியோரின் கூட்டுமுயற்சியில் உருவாக்கப்பட்டது. என்.எச்.எம். எழுதி வாயிலாக நமக்குத் தெரிந்த விசைப்பலகையைப் பயன்படுத்திக்கணினியில் உள்ளீடு செய்யலாம். என்.எச்.எம். எழுத்துருமாற்றி மூலம் பிற எழுத்துருக்களில் உள்ளவற்றை யூனிகோடு எழுத்துருவில் மாற்றிக்கொள்ளமுடியும்.

என்.எச்.எம் மைக்ரோசாப்ட் வோட் (Microsoft Word)இல் நேரடியாக, எளிமையாகப் பயன்படுத்தமுடியும். இதன் வாயிலாக Page maker இல் யூனிகோட் எழுத்துருவை மட்டும் பயன்படுத்தமுடியாது; ஏனைய எழுத்துருக்களைப் பயன்படுத்த முடியும்.

அளவில் சிறிய (88 KB – 9,11,029 bytes) என்.எச்.எம், இணையத்தில் இலவசமாகக் கிடைக்கிறது. இதன் காரணமாக இது உலக அளவில் பரவலான பயன்பாட்டைப் பெற்றிருக்கிறது.

முரசு அஞ்சல்

இச்செயலி முத்து நெடுமாறனால் உருவாக்கப்பட்டது. இது தமிழை உள்ளீடு செய்வதற்கான ஒரு செயலி ஆகும். இதன்மூலம் திஸ்கி, யுனிகோட் எழுத்துருக்களை உள்ளீடு செய்யலாம். எழுத்துருக்களை மாற்றுவதற்கு இது துணைபுரிகிறது. மேலும், பல தமிழ் வலைப்பக்கங்களில் நுழைந்து தமிழை வாசிக்கவும் உதவிபுரிகின்றது.

குறள் தமிழ்ச் செயலி

இச்செயலியில் பணிபுரிய ஒலிபெயர்ப்பு விசைப்பலகை, தமிழ் 99 விசைப்பலகை, புதிய மற்றும் பழைய தமிழ்த் தட்டச்சு விசைப்பலகை ஆகியவை பயன்படுகின்றன. மைக்ரோசாப்ட் விண்டோஸ் தொகுப்பில் இயங்கும் அனைத்து சாப்ட்வேர்களிலும் இதைப் பயன்படுத்தித் தமிழை உள்ளீடு செய்யமுடியும். அவ்வாறு செய்வதற்கு Unicode, TSC, TAB, TAM போன்ற எழுத்துருக்கள் உதவுகின்றன.

உயரிய தொழில் நுட்பத்துடன் தயாரிக்கப்பட்ட தமிழ் ஆங்கிலப் பயனர் இடைமுகம் இதில் உள்ளது. இது யுனிகோட் தொழில்நுட்பத்துடன் கூடிய தமிழ் ஆங்கிலச் சொற்செயலி ஆகும். மேலும் SMTP சார்ந்த மின்னஞ்சல் செயலியாகவும், ஜி மெயிலில் தமிழில் மின்னஞ்சல் செய்யவும் பயன்படுகிறது .எளிய நடையில் தமிழிலும் ஆங்கிலத்திலும் பயனர் கையேடு உள்ளது.

புதுவை தமிழ் எழுதி

இச்செயலியானது வலைப்பக்கத்துடன் ஒன்றிணைந்த ஒரு செயலியாகும். இதனை வந்தட்டில் சேமித்து வைத்தால் இணையத்தின் இணைப்பு இல்லாவிட்டாலும் பயன்படுத்தமுடியும். இதில் Bamini, Tscii, Tab, tam, Unicode ஆகிய எழுத்துருக்கள் செயல்படுகின்றன. இவ்வெழுத்துருக்களை அனைத்து இயங்குதளத்திலும் யுனிகோடுவழி உள்ளீடுசெய்து, மின்னஞ்சல்செய்யப் பயன்படுத்தலாம்.

துணைவன்

இது 1986இல் (MSDOS) மலேசியாவில் முதன்முதலில் வெளியிடப்பட்டது .இதனுடன் முதல் தமிழ் ஒலிவிசைப் பலகையும் (Tamil Phonetic Keyboard) வெளியிடப்பட்டது .துணைவனில் முதல் எழுத்துரு மாற்றி ,தன்னிச்சையான எழுத்துரு, குறியீடு, தமிழ்த் தரவுத்தளம் ஆகியன அமைந்துள்ளன .

சிங்கப்பூர், மலேசியா மற்றும் உலகளாவிய தமிழ்ப் பயன்பாட்டாளர்களின் மூலம் பல புதிய வளர்ச்சியை இது அடைந்தது .கல்வித்துறை மற்றும் வணிகம் சார் நிறுவனப் பயன்பாட்டிற்கு இது உதவுவதால், சிங்கப்பூர் மற்றும் மலேசியாவில் பல்லாயிரக்கணக்கான பயன்பாட்டாளர்களைக் கொண்டுள்ளது .இதில் அனைத்துத் தமிழ் விசைப்பலகையையும் பயன்படுத்தமுடியும் .உதாரணமாக ,தமிழ் ,99தமிழ்த் தட்டச்சு தரப்படுத்தப்படாத விசைப்பலகைகள் (Romanised Thunaivan, Mylai IE) போன்றவை அவற்றில் சிலவாகும் .TSCII, TAM, TAB and Unicode போன்ற அனைத்துத் தரப்படுத்தப்பட்ட குறியீடுகளையும்)Standard Encoding) இதில் பயன்படுத்த முடியும்.

இதன் புதிய பதிப்பான துணைவன் 7 என்னும் தமிழ் மென்பொருள், கணினியில் பயன்படுத்துவதற்கு எளிமையாகவும் தரத்தில் முதன்மையாகவும் இருக்கிறது. இந்த ஒரு மென்பொருள் மூலம் கணினியில் அனைத்து எழுத்துருக்களையும் பயன்படுத்தமுடியும். யுனிகோடு பயன்பாடு, பொறுப்புள்ள விசைப்பலகையில் நெகிழ்ச்சித் தன்மையுடன் தரப்படுத்தப்பட்ட பதிவிறக்கத் தன்மை, குறைந்த நேரத்தில் கணினியில் மென்பொருளை நிறுவுதல், எளிதில் doc files to Unicode and all 8 bit Tamil Encodings

மாற்றுதல் ஆகியவை சிறப்பாகும். மேம்படுத்தப்பட்ட நெகிழும் தன்மையுடைய எளிமையான உரிம முறைமையும் இதில் அடங்கியுள்ளது .

இதன் எளிமையான பயன்பாட்டிற்கு இணையாக வேறு மென்பொருள் எதுவும் சந்தையில் இல்லை . இதிலுள்ள எழுத்துரு மாற்றி ,பயன்பாட்டாளர்களுக்கு எளிதாக உள்ளது .இந்த எழுத்துரு மாற்றி)converter) எந்த ஒரு தரமான)standard) குறியீட்டையும்)encoding) மாற்றும் தன்மையுடையது. தமிழ் 99 Tam, Tab, Tscii, பழைய தட்டச்சு ,யுனிகோடு எனப் பல விசைப்பலகைகளை இது தன்னகத்தே கொண்டுள்ளது.

இளங்கோ தமிழ்

இது எம் எஸ் ஆபிஸ் ,மெயில் சாப்ட்வேர் ,பேஜ் மேக்கர் ,கோரல் ட்ரா ,போட்டோ ஷாப் ,ப்ளாஷ் போன்ற மென்பொருட்களில் தமிழை நேரடியாக உள்ளீடு செய்ய உதவுகிறது .

தமிழ் 99விசைப்பலகை ,தட்டச்சு விசைப்பலகை ,மரபார்ந்த ஒலியியல் விசைப்பலகை இவற்றை ஒருங்கிணைத்து இது செயல்படுகிறது .மேலும் ,ஆங்கிலத்தில் தட்டச்சுசெய்து தமிழில் பெறுவதற்கு ஏற்ற ஒலிபெயர்ப்பு விசைப்பலகையும் இதில் சேர்க்கப்பட்டிருக்கிறது .

எளிய பயன்பாடு மற்றும் பயனாளர் நோக்கில் இது வடிவமைக்கப்பட்டிருப்பதால் ,தேர்வு செய்த விசைப்பலகையில் குறைந்த விசை அழுத்ததின்மூலம்)key strokes) தமிழ் மற்றும் ஆங்கிலத்தைத் தொடர்ந்து பயன்படுத்த உதவுகிறது.

தமிழ் கீ

தமிழர்களிடையே பிரபலமாகியிருக்கும் விசைப்பலகைகளை இதன்மூலம் பயன்கொள்ள முடியும் .தற்போது அஞ்சல் ,தமிழ்நெட் ,பாமினி ,பழைய மற்றும் புதிய தட்டச்சு ,இன்ஸ்கிரிப்ட் மற்றும் அவ்வை விசைப்பலகை முதலியவற்றைப் பயன்படுத்த இது உதவுகிறது.

Alt+F6	- Avvai
Alt+F7	- Inscript
Alt+F8	- Anjal
Alt+F9	- Tamil 99
Alt+F10	- Bamini
Alt+F11	- Old Typewriter
Alt+F12	- New Typewriter

என மேற்குறித்த விசைப்பலகைகளை இதில் அமைத்துக்கொள்ள முடியும். மேலும், F9 விசையைக் கொண்டு ஆங்கிலம் ,தமிழ் இவற்றைத் தொடர்ந்து பயன்படுத்த இயலும் .

வானவில்

Vanavil Tamil 2000-DB-GEN.exe, Vanavil Tamil Interface.exe, Vanavil Vista-2008HP.exe, Vanavil W7-HP.exe, Vanavil-98-D2.exe, VANAVIL.EXE, Vanavil2000.exe, Vanavil98-6.0-D1.exe and VANAVILL.EXE. முதலிய வானவில் பதிப்புகள் கிடைக்கப் பெறுகின்றன .

அரும்பு

என்.எச்.எம். மென்பொருளைப் பின்பற்றி அரும்பு என்னும் தமிழ் மென்பொருள் உருவாக்கப்பட்டுள்ளது. தமிழில் பல்வேறு மென்பொருட்களில் உள்ளீடு செய்வதற்கான செயல்திறன் பெற்றது இது .இது இலவசமாகக் கிடைக்கிறது.

தொகுப்புரை

மென்பொருள்கள் மூலம் கணினியின் செயல்பாடு மிக எளிதாக மாறி வருவதை இக்கட்டுரை சுட்டிக்காட்டியுள்ளது .அத்துடன் மென்பொருள்களின் அவசியத்தையும் வலியுறுத்தியுள்ளது .இவ்வாறான தமிழ் மென்பொருட்கள் உருவாக்கத்தில் தமிழ்நாட்டில் உள்ள தமிழர்கள் மட்டுமன்றி மலேசியா, சிங்கப்பூர் ஆகிய நாடுகளில் வாழும் தமிழர்களின் பங்கும் இடம்பெற்றுள்ளமையைக் காணமுடிகிறது. இவை கணினித்துறையின் வளர்ச்சியை நமக்கு உணர்த்துவதாக அமைகிறது.

Information Retrieval System for Tamil and Non-Tamil Users

S.Srividhya

Student, M.C.A-CEG Anna University
Chennai-600 025
apssreevidhya@gmail.com

Dr.T.Mala

Senior Lecturer, Anna University
Chennai-600 025.
mala@cs.annauniv.edu

Abstract

This scheme enhances a remarkable approach to formulate and expand query. It accepts query given by user and expands it through a series of query expansion techniques. The input given by user is Tamil text and the output is expanded query to be given to a search engines to retrieve relevant English documents.

Keywords-Word sense disambiguation, Word stemming, Transliteration.

Introduction

The World Wide Web (WWW), a rich source of information is growing at an enormous rate. According to Online Computer Library Center, English is still the dominant language in the web that contributes most of the content. However, global internet usage statistics reveal that the number of non-English internet users is steadily on the rise, but all of them are not able to express their basic needs in English. Tamil users who are not able to express their needs in English are also growing in the Internet. They generally search for the information using the Tamil search engines. But the content provided by these search engines is less in number. Making the huge repository of information on the translation module takes the stemmed word web, which is available in English, accessible to non-English internet users has become an important challenge in recent times. When the non-English users want to access the existing search engines, most of the time they arrive at improper formulation of English queries. The proposed system aim to solve the above problem by allowing the users to pose the query in their own (source) language which is different from the language of the documents that are searched. This enables users to express their information need in their native language while the proposed system takes care of expanding the given user query that can be given to Search engines like Google, Alta vista.

LITERATURE SURVEY

“Query formulation for Information Retrieval System for Tamil Users” seeks to develop efficient techniques for query formulation whose accuracy and precision are determined by parameters like precision and recall. The query formulation process is broadly divided into three main portions: word stemming, translation and word sense disambiguation. The Word stemming is done with a morphological analyzer whose source code is modified to perform word stemming. The input is got in UNICODE and converted to TAB-Anna font . The translation module takes the stemmed word as input and transliterates those words to as input and transliterates those words to English and the equivalent

English words Bilingual dictionary is developed for this purpose. The word sense disambiguation makes use of Bootstrap algorithm [3] which disambiguates 55% of the words with 92% accuracy.

CLIR system for Agricultural society [8]: Tamil-English CLIR system for agricultural society uses Lesk algorithm for word sense disambiguation which is only 44% accurate and the Tamil words from the input query is directly searched in bilingual dictionary without checking for spelling mistakes.

Information retrieval system using mobile networks [5]: This system needs user interaction to disambiguate the word and the translation is prone to ambiguities if the context is not clearly specified by the user.

Cross Lingual Information Retrieval Using Data Mining Methods [6]: In this system the feedback and suggestions from the users are collected for document mapping. For each candidate word the pair wise measure gives a measure of correlation. However, these correlations are not available in dictionary representations and must be generated by use of appropriate ontological systems.

Tamil search engine[2]: This system discuss the issues related to crawler, database storage structure and other functional modules of the search engine. While they have shown some limited success, the approach used by the search engine is limited to the Tamil language and it retrieves the Tamil documents.

SYSTEM ARCHITECTURE

The system architecture is shown in Figure 3.1. The system mainly focuses on the construction of suitable English and the equivalent English words are fetched from a bilingual dictionary.. query for relevant English document retrieval in an information retrieval system. The proposed system gets Tamil input and retrieves relevant English documents according to the user query. The query is then given to spell checker. Morphological Analyzer obtains the root terms of source query by removing grammatical inflections. By applying rules for handling suffices, oblique, etc., the root words are obtained in the given query. Transliteration is done to convert Tamil letters to English characters in a systematic way. The output is then disambiguated to find out the exact equivalent English word. Word sense disambiguation is done that identifies the correct sense of an ambiguous word that is being used in a query.

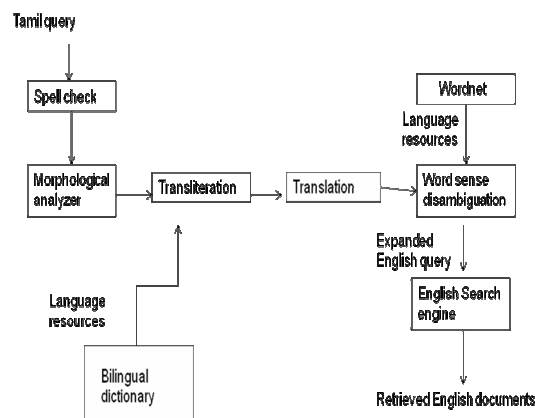


Figure 3.1 Word Sense Disambiguation

Translation is part of word sense disambiguation. For each sense of a given word, it is compared with all possible senses of the surrounding words in the given query and the word with maximum senses is chosen as the appropriate word. With the exact English words obtained as a result of word sense disambiguation, query formulation is done and given to search engines for English documents retrieval. If Tamil query is entered, the grammatical inflections are removed using morphological analyzer and query is formulated to retrieve relevant Tamil documents. WordNet, dictionary are the resources used. The formulated query is given to an existing search engine like Google, Alta vista. It uses bootstrap algorithm for Word sense disambiguation.

Bootstrap algorithm for Word sense disambiguation: Bootstrapping algorithm for Word Sense Disambiguation succeeds in disambiguating a subset of the words in the input text with very high precision. It uses WordNet as resource to disambiguate and for the purpose of identifying the correct sense of the words in a given text. The bootstrapping process initializes a set of ambiguous words with all the nouns and verbs in the text. It then applies various disambiguation procedures and builds a set of disambiguated words: new words are sense tagged based on their relation to the already disambiguated words, and then added to the set. This process allows us to identify, in the original text, a set of words which can be disambiguated with high precision; 55% of the verbs and nouns are disambiguated with an accuracy of 92%.

EVALUATION PARAMETERS

Precision

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search:

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Recall

Recall in Information Retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

For example for text search on a set of documents recall is the number of correct results divided by the number of results that should have been returned. The precision and recall values for five different queries have been calculated as follows.

Precision	0.72	0.89	0.78	0.76	0.80
Recall	0.77	0.86	0.84	0.78	0.82

Conclusion

The proposed System helps to find links of documents for health society and to retrieve the documents from a large corpus in English language. This system is very much useful for the users who can understand English but do not know to give appropriate English query. The users can pose their query in Tamil and retrieve relevant English documents by using this system. The Query formulation for information retrieval system for Tamil users displays the search result in English. It is appropriate, if the results are displayed in their own language for the users who do not know how to give query in English. The precision can be improved by employing Machine translation methods instead of word-by-word translation technique. The bilingual dictionary can be extended to include more words which will expand the scope of the system. This system can be further extended to Rank the pages and provide a summary (in English) of top pages, translate the summary to Tamil or provide an answer to the query in Tamil.

References

1. Anand kumar M, Dhanalakshmi V, Rajendran S, Soman K," A Novel Approach to Morphological Analysis for Tamil Language", Proceedings of the Tamil Internet Conference, 2009, pp.244 - 249.
2. Baskaran Sankaran."Tamil Search Engine", Tamil Internet, California, U.S.A, 2002.
3. Rada F.Mihalcea & Dan I. Moldovan," A highly accurate bootstrap algorithm for word sense disambiguation", International journal on artificial intelligence tools, 2001, Vol.10, No1-2.
4. M.B.A.Salai Aaviyamma and Dr.K.Kathiravan," Problems related to Eng-Tam Translation", Proceedings of the Tamil Internet Conference, 2009, pp. 169 - 172.
5. R.Shriram&Vijayan Sugumaran,"Cross Lingual Information Retrieval and Delivery Using Community Mobile Networks", IEEE, 2006.
6. R.Shriram&Vijayan Sugumaran,"Cross Lingual Information Retrieval Using Data Mining Methods", Americas Conference on Information Systems (AMCIS), 2009
7. Sinnathurai Srivas," Inside Tamil Unicode", Proceedings of the Tamil Internet Conference, 2009, pp. 140 - 144.
8. D.Thenmozhi and C. Aravindan, "Tamil-English Cross Lingual Information Retrieval System for Agriculture Society", Proceedings of the Tamil Internet Conference, 2009, pp. 173 - 178.
9. Dr. Vasu Renganathan," An Interactive Approach to Development of English-Tamil
10. Machine Translation System on the Web", Proceedings of the Tamil Internet Conference, California, U.S.A,2002

தமிழில் தேடுபொறி !

எம்.எஸ்.முரளி (செல்வமுரளி), த/பெ. செல்வராஜ்

தகவல் தொழில்நுட்ப ஆலோசகர்.

murali1309@gmail.com

அலைபேசி எண்: 99430-94945

தேடல் தளங்களும் அதன் பயன்களும்

வளர்ந்து வரும் நவீன உலகில் கணினி சார்ந்த தொழில்நுட்பங்கள் ஏற்படுத்திவரும் மாற்றங்கள் ஏராளம். அதிலும் குறிப்பாக இன்டர்நெட் எனப்படும் இணையத்தின் பயன்பாடு எல்லையில்லாமல் விரிந்து கொண்டே போகிறது. நினைத்த நேரத்தில் நினைத்த தகவல்களைப் பெறவும், தரவும், நினைத்தவுடன் பிறருடன் பேசவும் உதவுகின்றன. ஒருவரையொருவர் சந்திக்காமலேயே காரியங்களைக் கச்சிதமாக நிறைவேற்றும் வல்லமையை வகைப்படுத்தித் தந்திருக்கிறது இணையம். நவீனத் தொழில்நுட்ப வளர்ச்சியில் இணையத்தின் தேடல் தளங்கள் பெரும் பங்கு வகிக்கின்றன.

ஒட்டுமொத்த உலகத்தையும் உள்ளங்கைக்குள் சுருக்கி விட்டது இணையம். அதில் லட்சக்கணக்கான தளங்கள் இருக்கின்றன. என்ன தகவல்கள் தேவையென்றாலும் ஒரு சொடுக்கில் பெற்றுக் கொள்ளும் விதத்தில் தொழில் நுட்பம் விரல் நுனி வித்தையாகிப் போனது. இலட்சக் கணக்கான தளங்களின் பல கோடிக்கணக்கான பக்கங்களில் இருந்து நமக்குத் தேவையான தகவல்களை மட்டும் வடித்தெடுத்துத் தருவதில் தேடல் தளங்கள் பெரும் பங்காற்றுகின்றன.

தற்போது இணையத்தில் கூகிள் நிறுவனத்தின் கூகிள் (www.google.com), மைக்ரோசாப்டின் பிங் (www.bing.com), ஆகியவை பிரபலமான தேடல் தளங்கள். மேலும் மொழி சார்ந்த தேடலில் சீனாவின் பைடு (www.bidu.com) முன்னிலை வகிக்கிறது.

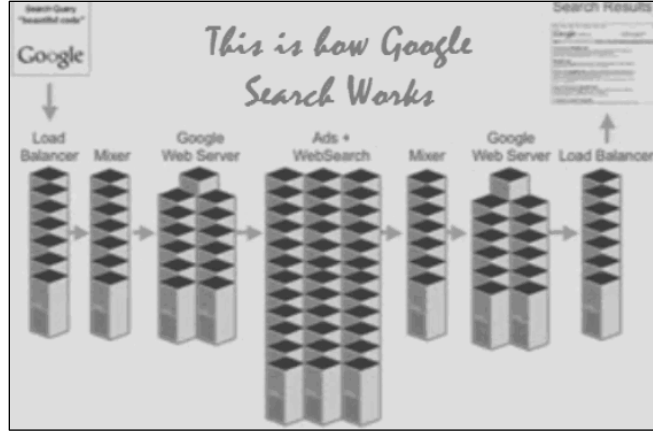
தற்போதுள்ள தேடல் தளங்கள் அத்தனைக்கும் முன்னோடி ஆர்சி (Archie) என்பதே. 'ஆர்கைவ்' என்பதன் சுருக்கமே இந்த 'ஆர்சி'. மெக்கில் பல்கலைக்கழகத்தின் 'ஆலன் எம்டாக்' என்ற மாணவர் கோப்புக்களை எளிதில் தேடும் வகையில் 'ஆர்சி' தேடற் பொறியை வடிவமைத்தார். இணைய வழங்கியின் ஆவணங்களை மேலாண்மை செய்யும் கோப்பு பரிமாற்று வரைமுறை (File Transfer Protocol - FTP) என்ற நுட்பத்தைப் பயன்படுத்தி கோப்பகப் பயன்பாட்டுப் பட்டியலில் (Directory Listing) குறிப்பிட்ட கோப்புகளைத் தேடும் வகையில் அவர் நிரலை (Programming) எழுதியிருந்தார். பின் தொழில் நுட்பம் வளர வளர அதற்கேற்றார்போல் பல தேடல் தளங்கள் வெளி வந்தன. இருப்பினும் 1998-ல் அறிமுகம் கண்ட கூகிள் நிறுவனத்தின் கூகிள் தேடல் தளம் பெரும் வரவேற்பை பெற்றது. அன்றிலிருந்து இன்று வரை கூகிள் நிறுவனம்தான் தேடல் தளச் சந்தையில் முன்னணி வகிக்கிறது.

தேடல் தளங்களின் வகைகள்.

தேடுபொறிகள் என்பது பொதுவாக **இணையத் தேடுபொறிகளையே** குறிக்கும். சில தேடுபொறிகள் உள்ளூர் வலையமைப்பில் மட்டுமே தேடும் விதத்தில் வடிவமைக்கப்பட்டிருக்கும். இன்னும் சில **தேடு பொறிகள் இணைய உலகில்** உள்ள பல நூறு கோடி பக்கங்களில் இருந்து நமக்குத் தேவையான தகவல்களில் மிகப் பொருத்தமான பக்கங்களைத் தேடித் தரும் ஆற்றலுடையவை. வேறுசில தேடற்பொறிகள் செய்திக் குழுக்கள், தகவல் தளங்கள், திறந்த இணையத் தளங்களைப் பட்டியலிடும் DMOZ.org போன்ற இணையத் தளங்களைத் தேடும்.

மனிதர்களால் எழுதப்பட்ட இணையத் தளங்களைப் பட்டியலிடும் தளங்களைப் போன்றல்லாது, தேடு பொறிகள் அல்காரிதங்களைப் பயன்படுத்தி தேடல்களை மேற்கொள்ளும். வேறு சில தேடற்பொறிகளோ தமது இடைமுகத்தை வழங்கினாலும் உண்மையில் வேறுசில தேடுபொறிகளில் தேடலை மேற்கொள்ளும். ஆரம்ப காலத்தில் ASCII முறை எழுத்துருக்களைக் கொண்டே தேடு சொற்களை உள்ளிட முடிந்தது. தற்போது ஒருங்குறி (Unicode) எழுத்து முறையைப் பல தேடுபொறிகளும் ஆதரிப்பதால் ஆங்கிலத்தில் மட்டுமல்லாது உலக மொழிகள் அனைத்திலும் மொழி சார்ந்த தேடல்களைப் பெற்றுக் கொள்ளும் வாய்ப்புக் கனிந்திருக்கிறது.

தேடல்தளங்கள் வேலை செய்யும் விதம்



தேடல்தளங்கள் மூன்று வகைகளாக வேலை செய்கின்றன

1. Web Crawler - இணையத்தளங்களில் தகவலைத் திரட்டுதல்

2. Indexing - திரட்டிய தகவலை உள்ளடக்குதல்

3. Searching - தேடுதல்

இணையத்தளங்களில் உள்ள தகவல்களைத் திரட்டச் சிலந்தி (Spider) என்றழைக்கப்படும் நிரல், இணையத்தளங்களின் உள்ளே முதலில் Robot.txt என்ற ஆவணத்தை நாடுகின்றது. அதில் அத்தளத்தின் தேடல் தளங்கள் எவைகளையெல்லாம் உள்ளடக்கலாம் என்ற விபரங்கள் இருக்கும். அந்த விபரங்களை மட்டும்தான் அந்த சிலந்திகள் தேடல் தளங்களுக்கு எடுத்துச் செல்ல வேண்டும். அந்தத் தளத்தில் உள்ள ஒவ்வொரு பக்க இணைப்புகளுக்கும் சென்று இணைப்புகளைச் சரிபார்த்துப் பின் அந்தந்த பக்கங்களின் தலைப்பு, மெட்டா டேக் , படங்கள் , படங்கள் பற்றிய விபரங்கள், காணொலி, காணொலி விபரங்கள் போன்றவற்றை தேடல் தளங்கள், தங்கள் தரவுத்தளங்களின் சேமித்துவைக்கின்றன. பின் இணையப் பயனாளர்கள் தேடல் தளங்களின் வழியே தேடும்போது தங்களது உள்ளடக்கத்தில் (index) அவ்வார்த்தைகளை சரிப்பார்த்துப் பின் தேடப்பட்ட வார்த்தைக்கு எது சரியான முடிவை தருகிறதோ அதை முதலாகவும், தேடிய வார்த்தைகளோடு தொடர்புடைய வார்த்தைகளை அதன் பின்னரும் வரிசைப்படுத்திக் காட்டும்.

தமிழில் தேடல் தளம் தேவைதானா?

இதுவரை இணையத்தில் பல்வேறு புதிய புதிய தேடல்தளங்கள் வந்தாலும் அவையனைத்தும் ஆங்கிலம் மற்றும் இதர மொழிகளில் செயல்படும் விதத்தில் அமைந்துள்ளன. ஆனால் தமிழ் என்று வரும்போது சில தொழில்நுட்பச் சிக்கல்களால் துல்லியமான முடிவுகள் கிடைப்பதில்லை.

கூகிள் , பிங், யாஹூ, குயில் என பல தேடு பொறிகளில் கூகிள் சிறந்ததாக விளங்கி வந்தாலும் தமிழ் வழித் தேடல் என வரும்போது நாம் எதிர்பார்த்த அல்லது தேடிய வார்த்தைகள் அகப்படுகின்றனவா? நிச்சயமாக நம்முடைய தேடல் முழுமையாகக் கிடைக்கவில்லை என்பதே பதில். எனவே முழுவதும் தமிழுக்கென்று தனித்துவமான ஒரு தேடல்பொறி அவசரத் தேவை.

தமிழின் பாரம்பரியம் மற்றும் கலாசாரம் என்று ஆங்கிலம் வழியாக தேடும்போது மற்ற ஆங்கில நாடுகளின் கலாசாரப்பகுதிகள்தான் முதன் முதலில் புலனாகின்றன. உதாரணமாக: Tamil Culture என்று தேடினால் விக்கிபீடியா மற்றும் இதரத் தளங்களில் உள்ள தகவல்களை மட்டுமே எடுத்துத் தருகிறது; ஆனால் தமிழ்ப் பண்பாடு பற்றி முழுமையாக எழுதியிருக்கும் வலைப்பூக்கள் மற்றும் இணையத்தளங்களில் உள்ள தகவல்கள் கடைசியில்தான் இடம் பெறுகின்றன. இதனால் சில நேரங்களில் நமக்குத் தேவையான தகவல்கள் இணையத்தில் இருந்து கிடைக்காமலும் போய்விடுகின்றன. மேலும் "என்னவென்று, மறைந்திருந்து" என்பன போன்ற சொற்களைக் கொடுத்துத் தேடும்போது "என்னவென்று" இருக்கும் தகவல்களை மட்டும் காட்டுகிறது. ஆனால் "என்ன என்று", "மறைந்து இருந்து" என்ற வரும் தகவல்கள் அகப்படுவதில்லை.

தமிழ் தேடல்தளம் எப்படி இருக்கவேண்டும்?

இன்றைய காலகட்டத்தில் இணையத் தேடல் தளங்களில் மிக அதிகமான பயனாளர்களைப் பெற்றிருப்பது கூகிள் தேடல் பொறி என்று ஓர் ஆய்வு சொல்கிறது. அதனால்தான் இணையப் பயனாளர்கள் அந்தத் தேடல் தளத்தினை 'கூகிளாண்டவர்' என்று அழைக்கின்றனர். கூகிள் நிறுவனத்தின் வெற்றிக்குரிய முக்கியக் காரணிகளில் அந்தத் தேடல் தளத்துக்குப் பெரும் பங்கு உண்டு.

ஓர் இணையத் தளத்தினைத் தன்னுள்ளே உள்ளடக்கும்போது அந்தத் தளத்தின் முழுப்பக்கங்களையும் கூகிள் தரவுத்தளத்தில் அட்டவணைப்படுத்திவிடும். ஆகையால் கூகிளைப் பொறுத்தவரை எப்படித் தேடினாலும் சரியான பொருளை எடுத்துத் தந்துவிடும். இப்படி அடுக்கிக் கொண்டே போக இன்னும் பல சிறப்பம்சங்கள் கூகிளில் உண்டு. அதைப்போலவே நம் அன்னைத் தமிழிலும் ஒரு தேடல்தளம் வர வேண்டும். அவற்றில் பின்வரும் அம்சங்களும் இருந்தால் சாலச்சிறப்பு.

- தானியங்கு ஆலோசனை (Auto Suggest)
- தானியங்கு நிறைவி (Auto Complete)
- துறை சார்ந்த தேடல் (Category)
- உள்ளிணைந்த தமிழ் தட்டச்சும் வசதி (Tamil Input features)
- புத்தக தேடல் (Book Search)
- காணொலி (Video Search)

தமிழ் இலக்கண ரீதியான சிக்கல்கள்

- தமிழுக்கென்று வரும்போது தமிழ் இலக்கணத்தை இங்கே அடிப்படையாக வைத்து உருவாக்கினால் மட்டுமே தமிழ்த் தேடல் முழுமைபெறும்.
- தமிழ் தேடல் தளங்களில் தமிழில் தேட சில சமயங்களில் வார்த்தைகளைப் பிரித்துத் தேடாமல் இருக்க வேண்டும். எடுத்துக்காட்டாக 'பசுபதி' என்று தேடும்போது 'பசு'வும் 'பதியும் பிரித்து இடக் கூடாது.
- அதேபோல், தமிழில் இருக்கும் வார்த்தையைத் தேடும்போது அதன் ஆங்கில அர்த்தத்தைக் கொண்டு தேடினாலும் சில சமயங்களில் நன்மை பயக்கும். அடுத்ததாக ஆங்கிலத்தில் தேடினாலும், அதைத் தமிழில் தேடிக் கொடுக்கவேண்டும்.
- துறை சார்ந்த தேடல்களாக இருக்கவேண்டும்.
- எழுத்துப்பிழைகளைத் தவிர்த்துத் தேடுவது சிறப்பு. (உம்) அண்ணன் என்று தேடுவதற்குப் பதிலாக அண்ணன் என்று தேடினாலும் ஒரே பொருள் வருமாறு இருக்கவேண்டும்.

- அதேபோல் வார்த்தைகளின் உள்ளே வரும் இடைவெளிகளையும் கவனிக்க வேண்டும்.(உம்) 'தேடல்' ற்கும், 'தே டல்' ற்கும். இங்கே இருக்கும் வெற்றிடங்களை கவனித்து ஒரே பொருளைத் தரவேண்டும்.

தொழில்நுட்ப ரீதியான சிக்கல்கள்

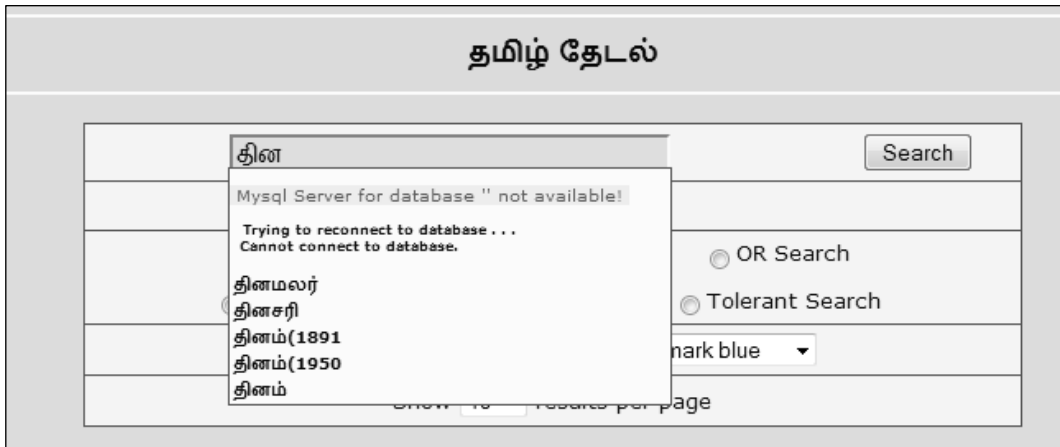
1. இது போன்ற அதிசக்தி வாய்ந்த நிரல்களை பகிர்ந்தளிக்கப்பட்ட வழங்கியில் (Shared Hosting) பயன்படுத்தும்போது அனுமதிக்கப்பட்ட நினைவகம் மற்றும் நிரல்கள் இயங்கும் நேரம் ஆகியவை நிபந்தனைக்குட்பட்டது. எனவே ஒரு தளத்தில் உள்ள தகவல்களைத் திரட்டும்போது அனுமதிக்கப்பட்ட நேரத்திற்குப் பின் தகவல் திரட்டுவது நிறுத்தப்படும். எனவே முழுமையான தேடல் இவ்விடத்தில் நிறைவேறாது.
2. மேலும் புத்தகத்தேடல், காணொலி தேடல் போன்றவற்றை உருவாக்கினாலும் அவற்றை நமது வழங்கியில் சேமித்திட அதிகமான இடம் உள்ள வழங்கி தேவை. கூகிள் நிறுவனம் தனது தேடல் தளங்களுக்கு மட்டும் 6000 இணைய வழங்கிகளை வைத்துள்ளது குறிப்பிடத்தக்கது. மேலும் வருடந்தோறும் இணைய வழங்கிகளையும், தகவல் சேமிப்பகத்தையும் அதிப்படுத்திவருவது குறிப்பிடத்தக்கது.
3. எனவே இதுபோன்ற பயன்பாடுகளுக்கு நேரடியாகத் தனித்தியங்கும் வழங்கிகள் (Dedicated server) தேவை. ஆனால் இந்த வழங்கிகளின் பயன்பாட்டுக் கட்டணம் அதிகம். ஆனால் அதன் வழியாக ஒரு முழுமையான நேரடியான தமிழ் தேடல்தளத்தினை அளிக்க முடியும்.

தொழில்நுட்ப மென்பொருட்கள்

இன்று இணையத்தள மென்பொருட்கள் உருவாக்கத்தில் மைக்ரோசாப்ட் நிறுவனத்தின் ஏஎஸ்பி (ASP) மற்றும் கட்டற்ற மென்பொருட்களான PHP, PYTHON, PERL போன்றவையும், ஜேஎஸ்பி (JSP) போன்றவையும் புகழ்பெற்று விளங்குகின்றன. லினக்ஸ் இயங்குதளத்தில் Php / Mysql / Java Script ஆகியவற்றை அடிப்படையாக வைத்து இத்தமிழ்த் தேடல்தளம் உருவாக்கப்பட்டுள்ளது. மேலும் தானியங்கு நிறைவு (Auto Complete) வசதிக்காக பிரத்யேகமாக ஒரு தனி ஜாவாஸ்கிரிப்ட் அடிப்படையாக வைத்து நிரல் எழுதப்பட்டுள்ளது.

மேலும் தமிழ் தேடல்தளத்தில் உள்ளிணைந்த தமிழ் தட்டச்சும் வசதிக்காக www.higopi.com ல் உள்ள தகடுர் நிரல் பயன்படுத்தப்பட்டுள்ளது.

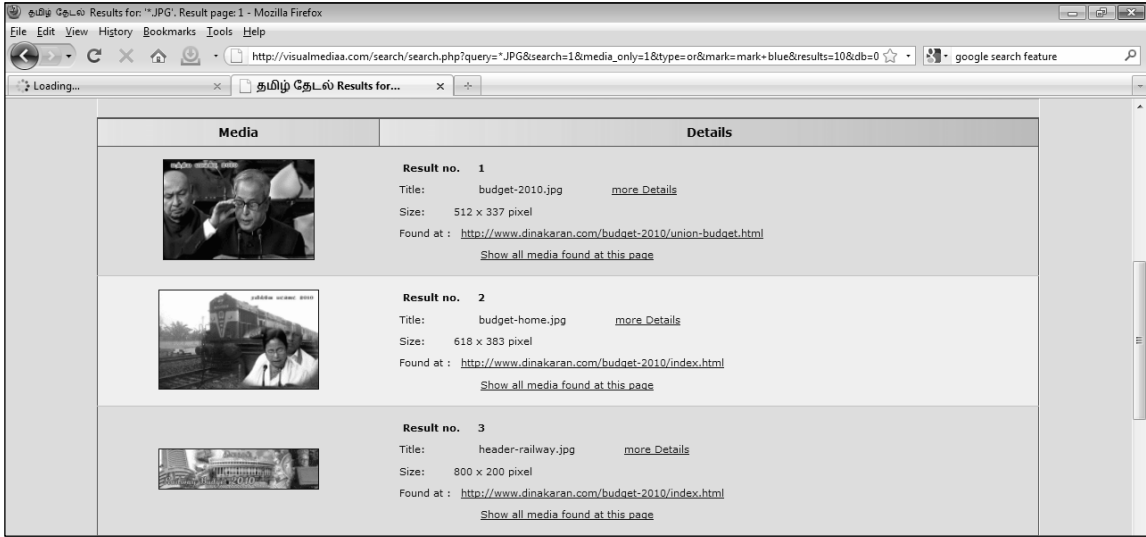
படம்1 : தானியங்கு நிறைவி (Auto Complete)



படம் 2: வகைப்படுத்திய தேடல்



படம் 3 : படத் தேடல்



முடிவுரை :

இந்தத் தமிழ் தேடல்பொறியை விஷுவல் மீடியா நிறுவனம் (www.visualmediaa.com) உருவாக்கிக் கடந்த ஒன்றரை ஆண்டுகளாகச் சோதனை செய்து வருகிறது. ஆரம்பத்தில் இத்தமிழ் தேடல்பொறி சொற்றொடர்களை மட்டுமே தேடும்படி அமைந்திருந்தது. தற்போது தமிழ்ப்பெயரில் உள்ள படங்கள், காணொலி மற்றும் பிடிஎஃப்(PDF) கோப்புகளையும் தேடித்தருமாறு வடிவமைக்கப்பட்டிருக்கிறது. இந்த தமிழ் தேடல்பொறியை மேலும் சிறப்பாக இயங்கச் செய்திட தனித்தியங்கும் வழங்கியும் (Dedicated Server), இன்னும் சில தமிழ் அறிஞர்கள், பொறிஞர்கள் அடங்கிய குழுவும் அமையப்பெற்றால் நிச்சயமாக இன்னமும் ஒரு வருடத்திற்குள் முழுமைபெற்ற தமிழ் தேடல் தளத்தினை வெளியிட இயலும்.

Searchko - The King of Search for Tamil Web Documents

Sobha Lalitha Devi, Pattabhi R K Rao T, Vijay Sundar Ram R

Au-KBC Research Centre, MIT campus of Anna University, Chennai- 600 044

{sobha, pattabhi, sundar}@au-kbc.org

Abstract

Searchko is a Tamil portal, which uses information retrieval (IR) technology for searching the Tamil content in the web. The Searchko engine uses many types of natural language processing technology for getting the most relevant output. The web has nearly 10 million documents in Tamil and bringing it under one umbrella using IR technology is the aim of Searchko. The etymology of searchko is Search+Ko ie search+ king, ko in Tamil means king and thus this portal is a unique portal where you can get all the content , whether it is News from the news papers, Cinema, Tamil literature, Music and Cricket. The unique features of SearchKo are Multiple Font support, Enhancing the query with the help of a morph analyzer, Works with a phonetic visual keyboard, Spell Checking, Query expansion using thesaurus and dictionaries.

Introduction

The growth of technology and internet has brought information revolution in our country and across the world. This 21st century is called as information age. This has changed the way people share the knowledge, do business, and interact with each other. Until 1990's internet was dominated by only English content. Today the World Wide Web has grown wider and has become very large, having content in all Indian languages. And especially in Tamil, web has more than 10 million documents. With this huge amount of Tamil data available on web, we require systems which will enable users to easily search and access data. In the present paper we present a description of Searchko, a Tamil portal, whose objective is to provide Tamil users access to all Tamil content on web.

Searchko - an overview

Searchko provides various contents for the users. The search contents are classified into different domains such as Literature , Health , and Cinema search. The content for health domain is created in-house from the health texts. The results obtained for health search are classified into allopathy, siddha, homeopathy, ayurveda, when presented to the user. For example users interested in reading articles related to cancer, can give query as "puRRu noy" and obtain documents from ayurveda, allopathy, homeopathy, siddha. In the literature search, we can have focused search on contents of Tamil literatures of wide range from sangam literature to 20th century creations. This includes 'aimperum kaappiyam', 'ettuth thokai', devotional literatures such as 'kuravanji paataalkal', 'pakthi ilakkiyangkal', text having poems by Sidhars such as 'sidhar paataalkal' and present days kalki and jayakanthan stories. The content for general search consists news articles from online news magazines, Tamil wikipedia, blog sites and other Tamil sites.

All these contents are crawled and indexed by search engine regularly and updated periodically. This process is fully automated. It has been observed that even though we have huge content in web in Tamil, people are not able to have access for these documents because most of the content is not available for search. By developing this portal we have made all the Tamil content on web accessible to the users.

Another important feature is the lexical resource most often referred by people of all age groups as dictionaries is available for search and there are English – Tamil, Tamil – English dictionaries. We have found that there are different online dictionaries available in the web, but most of these do not provide the exact meaning of words in all senses. For example for the word “bank”, There are different noun senses and verb senses such as “financial institution”, “river bank”, “to take support on someone or something”. Searchko provides online dictionaries, where all senses of the word are given along with the word’s part-of-speech category. The user interface of the dictionaries is very simple and user friendly. English – Tamil dictionary consists of more than 150000 root words. And similarly Tamil – English dictionary consists of more than 100000 words. The Tamil – English dictionary also includes old Tamil words, for which finding meanings in today’s new dictionaries is quiet difficult. The English – Tamil dictionary includes Technical terms also. These dictionaries were created by lexicographers and verified by linguistic experts. These are very helpful to Translators.

Searchko provides entertainment contents focusing cinema, music and sports. In Cinema section, we provide details of actors, film directors, music directors, playback singers, and producers for each Tamil film. The Cinema database is updated whenever a new film is released. Users can search by giving actors names or film names or music director’s names or producer’s names. For example if the user searches for actor “Rajnikanth”, the Cinema search provides all films acted by Rajnikanth, for each film of his, will be provided with details of producer’s name, director’s name, playback singers name, Date of film release and also will fetch a relevant video of the film from You tube, if it exists. This cinema section also provides users with list of new upcoming films, their promotional videos if it exists. This can be considered as Tamil IMDB.

The music section of the portal is focused on music festival of “markazhi” month (Dec –Jan) also known “markazhi thiruvizhaa” at Chennai. Here we provide schedules of all music events that take place in various auditoriums or “sabhas” in and around Chennai. The schedules can be searched based on date or artist name or auditorium name or by time of event. The music lovers of the city have found this very helpful as they can find the time and venue of their interested event. This section also provides devotional hymns of Tiruppavai, Tiruvembavai and Tirupalliezuchi along with their meanings in Tamil. These devotional hymns are sung commonly during this markazhi month in various temples and households of Tamil Nadu.

In the Searchko, we provide live cricket scorecard in Tamil. The Cricket scorecard in English is obtained from websites such as cricinfo, cricbuzz, willow, and these scores are translated on-line to Tamil. The translation is done automatically. Here we use template based extraction and translation. For translating names of players, we use transliteration engine. The transliteration engine is built using statistical methodology [1]. The engine is trained using a parallel named entity lexicon. The engine works with an accuracy of 93%. The content in the whole web portal, is fetched and displayed automatically without any manual intervention.

The Natural language Processing modules of Searchko

Searchko has many unique features, which are developed using sophisticated natural language processing techniques.

a) Multiple Font Encoding Support:

Tamil content on web are in various encoding schemes and not as in English. There is a large data available in proprietary fonts. On giving a query the system searches all the Tamil pages independent (covering most of the proprietary fonts) of fonts to get the relevant pages. And snippets for all the retrieved documents are given in Unicode. To achieve this font independency in the search engine, we have used different transcoder engines to unify the contents in to Unicode. A font transcoder engine identifies the encoding scheme of one font and makes an equivalent map to another font encoding scheme. Here we have converted all proprietary fonts into UTF-8 encoding scheme. For the purpose of creating equivalent map, both encoding schemes have to be analysed and each glyph of proprietary font has to be mapped to the Unicode scheme. There can be one to many or many to one mapping.

b) Enhancing the query with the help of a morph analyzer:

Tamil being a morphologically rich language, multiple words can be generated from a given root word. For eg the word “padi” will have words like “padithaan”, “padithithaal” etc. Here in our search engine, we enhance the query with the help of a morphological analyzer to retrieve all documents which have the various forms of the given query word.

A morphological analyzer is a language processing tool, which will segment a given word into root word and its suffixes and will give their syntactic information respectively. .

Example ‘viitukal’ -> ‘viitu’ + ‘kal’

Here in this task, the morphological analyzer is used only to get the root word. This is done using Finite State Automata and paradigm based approach [4]. This works with an average precision of 92.13%

For example if the query word is ‘malarkal’ the search engine will look for documents having words such as ‘malar’, ‘malarkalai’, ‘malarai’, etc. This enables the user to give a query in crisp and easy form and get retrieved documents having all possible forms of the query word.

c) Query expansion using thesaurus and dictionaries:

We use a Tamil thesaurus and dictionary to retrieve more documents with the same sense. This increases the relevant documents [2]. For example, for the query word ‘puu’, documents are searched for other words with same sense like ‘malar’, ‘pushpam’, ‘alar’, ‘koNtai’, ‘sutaan’ along with the given query word ‘puu’.

The thesaurus is an advanced form of a lexicon, which contains the root words and their synsets in contrast to a lexicon which contains meaning and pronunciation of words. Here we have used an electronic thesaurus, which contains synsets of the root words.

The thesaurus and the dictionary contain around 75000 words. This resource is increased as and when we encounter new words. Here the search handles AND, OR and Quoted queries.

navigate the web in Tamil, their on mother tongue. In this advance language processing technologies have been used to obtain good results. The search is available at <http://www.searchko.in>.

Reference

1. Mohammad Afraz And Sobha L (2008), 'English To Dravidian Language Machine Transliteration: A Statistical Approach Based On N-Grams', In The Proceedings Of International Seminar On Malayalam And Globalization, Trivandrum, Kerla.
2. Pattabhi R. K. Rao and Sobha L (2008), "AU-KBC FIRE2008 Submission - Cross Lingual Information Retrieval Track: Tamil-English", First Workshop of the Forum for Information Retrieval Evaluation (FIRE), Kolkata. pp 1-5.
3. Vijay Sundar Ram R., Chandra Mouli N., Bhuvaneswari P., Ananda Priya J. and Kumara Shanmugam B. (2005), "Hybrid Approach for Developing a Tamil Spell Checker", In the Proceedings of International Conference on Natural Language, Indian Institute of Technology, Kanpur, pp. 111-115.
4. Vijay Sundar Ram R., Menaka S., and Sobha Lalitha Devi. (2010). "Tamil Morphological Analyser", In the Proceedings of Knowledge Sharing Event on Morphological Analysers and Generators, LDC-IL, CIIL, Mysore. pp 1-18.

கோரி தேடல் அமைப்பு

பொருள் விரிவாக்க முறை

இளஞ்செழியன், கீதா, ரஞ்சனி பார்த்தசாரதி, மதன் கார்க்கி

கட்டுரைச்சுருக்கம்

தற்போதுள்ள இணைய தள தகவல் தேடல்கள், சொல் சார்ந்த தேடல்களாகவே இருந்து வருகின்றன. சொல் சார்ந்த தேடலின் போது சொல்லுக்குரிய கோப்பு கிடைக்காவிடில் அச்சொல்லின் பொருளுடைய வேறு சொற்கள் சார்ந்த கோப்புகளை பெறுவது இயலாது. இது தமிழ் தகவல் தேடலில் ஒரு குறைபாடாகவே இருந்து வருகின்றது. இக்குறைபாட்டை நீக்க பொருள் சார்ந்த தேடல் முறை 'கோரி' என்னும் இணைய தேடல் மூலம் அறிமுகம் செய்யப்பட்டுள்ளது. இணைய தேடலில் சொல்விரிவாக்கம் ஒரு அங்கமாகும். 'சொல் விரிவாக்கம்' என்பது தேடப்படும் சொல்லோடு தொடர்புடைய சொற்களை விரிவுபடுத்தும் செயல் ஆகும். இதன்மூலம் சொல் விரிவாக்கங்களை மட்டுமே அடிப்படையாகக் கொண்டு கோப்புகள் பெறப்படுகின்றன. இதனால் அதே பொருளில் இருக்கும் மற்ற கோப்புகள் தேடல் முடிவுகளில் வருவதில்லை. இந்தக் கட்டுரையில் 'பொருள் விரிவாக்கம்' எனும் புதிய முறை அறிமுகப்படுத்தப்பட்டுள்ளது. இந்த விரிவாக்க முறை, தேடப்படும் சொல்லை அதன் பொருளுடன் சேர்த்து விரிவாக்கம் செய்கிறது. சொல்லின் விரிவாக்கங்களையும் அதன் பொருளின் விரிவாக்கங்களையும் கொண்டு தேடப்படும் தகவல்கள், 'சொல் விரிவாக்கம்' முறையில் கிடைக்காத கோப்புகளையும் தேடுவதற்கான உள்ளீடுகளைத் தருகிறது. இதன் மூலம் பொருத்தமான முடிவுகளை துல்லியமாகப் பெற முடியும். பொருள் விரிவாக்கங்களுக்கு, 'உலக இணைய மொழி' (உ.இ.மொ) (Universal Networking Language)[1] என்ற இடைநிலை வகைக்குறி முன்மொழியப்படுகிறது. கோரி என்ற தேடல் அமைப்பின் ஒரு அங்கமாக இந்த 'பொருள் விரிவாக்கம்' அமைகிறது. உ.இ.மொ விதிகளின் படி இந்த 'பொருள் விரிவாக்கம்' நிகழ்வதால், 'கோரி' அமைப்பின் 'பொருள் சார்ந்த அடுக்குக்குறியில் தேடுவது' எளிமைப்படுத்தப்படுகிறது. இந்தப் 'பொருள் விரிவாக்க' முறையால் முடிவுகளின் துல்லியம் அதிகரித்திருப்பது இந்தக் கட்டுரையின் முடிவுகள் பகுதியில் ஆதாரத்துடன் நிரூபிக்கப்பட்டுள்ளது.

முன்னுரை

இணைய தள தகவல் தேடல் சொல் சார்ந்த தேடல்களாகவே இருந்துவருகின்றன. சொல் சார்ந்த தேடலின் போது சொல்லுக்குரிய கோப்பு கிடைக்காவிடில் அச்சொல்லின் பொருளுடைய வேறு சொற்கள் சார்ந்த கோப்புகள் பெறுவது இயலாத நிலையாக உள்ளது. இந்த குறைபாட்டை நீக்க 'பொருள் தேடல்' என்னும் புதிய முறை அறிமுகம் செய்துள்ளோம். பொருள் தேடலுக்கு உலக இணைய மொழி (உ.இ.மொ) என்னும் இடைநிலை வகைக்குறி முன்மொழியப்படுகிறது. உலக இணைய மொழி என்பது மொழிகளுக்கு இடையே உள்ள தகவல் மற்றும் ஆய்வறிவுகளை உரிய முறையில் பெற உதவுகிறது. உலக இணைய மொழியானது தகவல் மற்றும் ஆய்வறிவுகளை சொற்பொருள் பிணையமாக (semantic network) அமைந்திருக்கும். உலக இணைய மொழியானது உலக சொல் (Universal Word), தொடர்பு நிலை (Relation), ஏற்பி (Attribute) மற்றும் உ.இ.மொ வின் ஆய்வறிவு ஆதாரங்களை (Knowledge Base) உள்ளடக்கியுள்ளது. உலக சொல் என்பது உ.இ.மொழியின் சொல்லகராதியாகும். தொடர்பு நிலை (உறவு நிலை) மற்றும் ஏற்பி (பண்பு) உ.இ.மொழியின் இலக்கண விதிகளை உள்ளடக்கியது. உ.இ.மொ அறிவுத்தளம் என்பது உ.இ.மொழியின் சொற்பொருள் ஆய்வியலாகும்[1].

சொல் சார்ந்த பொருள் தேடல் என்பது இணையத்தில் குவிந்த பல பில்லியன் பக்கங்களில் இருந்து சொல்லுக்கு பொருத்தமான பக்கங்கள் மட்டுமின்றி சொல் சார்ந்த பொருளுடைய பக்கங்களை தேடித்தருவதற்கு உதவுகின்றது. இந்தத் தேடுபொறி இணைய தளத்தில் 'மறைமுக' (Offline) மற்றும் 'நேரடி' (Online) என்று இரண்டு நிலைகள் செயல்படுகிறது. 'மறைமுக' நிலையில் இணைய தள பக்கங்கள் 'இணையத் தவழி' (Web Crawler) மூலமாக பெறப்படுகிறது. பின்னர் பெறப்படும் பக்கங்களில் உள்ள வாக்கியங்களை பிரித்து, அவ்வாக்கியத்திலிருக்கும் தகவல்கள் ஒரு சுட்டு ஆக்க வரிசையில் சேமிக்கப்படுகின்றன. பின்னர் சில நெறி முறைகளை பயன்படுத்தி சொல் விரிவாக்க உள்ளீடுகள் தகவல் தளத்தில் சேகரிக்கப்படுகின்றன. 'நேரடிநிலையில்' பயனர் கோரும் சொல்லுக்கு பொருத்தமான கோப்புக்களையும், பொருள் விரிவாக்கத்திற்கான கோப்புகளையும் சுட்டு ஆக்க வரிசையின் மூலம் பெறப்பட்டு, அக்கோப்புகளை தரவரிசைபடுத்தி (Ranking) பொருத்தமான முடிவுகளைக் பெறமுடியும். இந்த இணைய தகவல் தேடல் சுற்றுலா சம்மந்தமான தேடுபொறி ஆகும்.

இணைய தள தேடலில் சொல் விரிவாக்கம் ஒரு அங்கமாகும். சொல் விரிவாக்கம் ஒரு 'மறைமுக' (Offline) முறையாகும். கூகிள், பிங் போன்ற இணைய தேடல்கள் பல சொல் விரிவாக்க வழிமுறைகள் பயன்படுத்தப்படுகிறது[3]. இவ்வகையான வழிமுறைகள் மூலம் சொல் அதிர்வெண்கள் கொண்டு கோப்புகள் தொகுக்கப்பட்டுள்ளது. அத்தொகுப்பில் உள்ள மிகுதியான அதிர்வெண்ணை கொண்ட சொல்லுக்கு இணை நிகழ்வுச் சொல் (co occurrence word) எடுக்கப்படுகிறது. பின்னர் ஒரு சொல்லுக்கு மிகுதியான அதிர்வெண்கள் உடைய இணை நிகழ்வுச் சொல்லைக் கொண்டு சொல் விரிவாக்கம் செய்யப்படுகிறது. இவ்வகையான சொல் விரிவாக்கத்தின் மூலம் சொல் சார்ந்த பொருள் விரிவாக்கம் செய்ய இயலாத நிலை உள்ளது. இந்த நிலையை போக்க 'பொருள் விரிவாக்கம்' என்னும் புதிய முறை அறிமுகப்படுத்தப்பட்டுள்ளது.

இந்த ஆய்வுரை கட்டுரையில் 'பொருள் விரிவாக்கம்' என்னும் புதிய முறை அறிமுகப்படுத்தப்பட்டுள்ளது. இந்த விரிவாக்க முறை, தேடப்படும் சொல் அதன் பொருளுடன் சேர்த்து விரிவாக்கம் செய்கிறது. சொல்லின் விரிவாக்கங்களையும் அதன் பொருளின் விரிவாக்கங்களையும் கொண்டு தேடப்படும் முறை ஆகும். பொருள் விரிவாக்கங்களுக்கு, 'உலக இணைய மொழி' (உ.இ.மொ) (Universal Networking Language) என்ற இடைநிலை வகைக்குறி முன்மொழியப்படுகிறது. 'இணையத் தவழி' மூலமாகப் பெறப்படும் கோப்புகள் உ.இ.மொ விதிகளுக்கு உட்படுத்தி பெறப்படும் தகவல்கள் அடுக்குக்குறி வடிவூட்டம் செய்யப்படுகிறது. சுட்டு ஆக்க வரிசை சார்ந்த 'பொருள் விரிவாக்கம்' முறையால் சொல் மற்றும் சொல்லின் பொருளுக்குபொருத்தமான கோப்புகள் பெறமுடியும். இதன் மூலமாக தேடுபொறியின் துல்லியம் அதிகரிக்கும். இந்த சொல் விரிவாக்கம் கோரி என்னும் தேடல் அமைப்பின் மூலம் சோதனை செய்யப்பட்டுள்ளது.

பின்புலம்

இணைய தேடல்களில் பல சொல் விரிவாக்க வழிமுறைகள் உபயோகப் படுத்தப்படுகிறது. அவ்வழிமுறைகளில் இணை நிகழ்வு அகராதி அடிப்படையிலான சொல் விரிவாக்க (Co occurrence-thesaurus-based expansion) முறையில் ஒரு கோப்பில் மிகுதியாக அமைந்திருக்கும் இணைச் சொல்லுக்கான சொற்பொருள் மட்டும் அதன் தொடர்பு நிலையைக் கொண்டு சொல் விரிவாக்கம் செய்யப்படுகிறது[8]. சொற்பொருள்களின் தொடக்கம் சார்ந்த சொல் விரிவாக்கத்தில் (Ontological Query Expansion) ஒரு சொல்லுக்கான பொருள் மற்றும் 'இன்' அல்லது 'உடைய' (possessor of) என்னும் தொடர்பு நிலைகள் கொண்டு சொல் விரிவாக்கம் செய்யப்படுகிறது[6]. உதாரணத்திற்கு "சிறகுகளுடைய பறவை" அல்லது "பறவையின் சிறகுகள்". பலமொழித் தகவல் தேடலில் (தமிழ் <-> ஆங்கிலம்), ஒரு சொல்லுக்கு இணையான சொற்களை 'சொல் தரவுத் தளத்தில்' இருந்து (Lexical database or WordNet)[9,10] பெறப்பட்டு சொல் விரிவாக்கம் செய்யப்படுகிறது[7]. அக்லோமேரடிவே தொகுதி[11] (Agglomerative Clustering)

வழிமுறையில் இரு கோப்புகளுக்கு கொசைன்(cosine) சமானம் மூலம் பல தொகுதிகளாக தொகுப்பதாகும். அத்தொகுதியில் உள்ள கோப்புகளில் மிகுதியான அதிர்வெண்கள் கொண்ட சொல்லுக்கு மிகுதியான அதிர்வெண்கள் கொண்ட இணை நிகழ்வுச் சொல்லை கொண்டு சொல்விரிவாக்கம் செய்யப்படுகிறது.

சொல் மற்றும் பொருள் விரிவாக்க முறையியல்

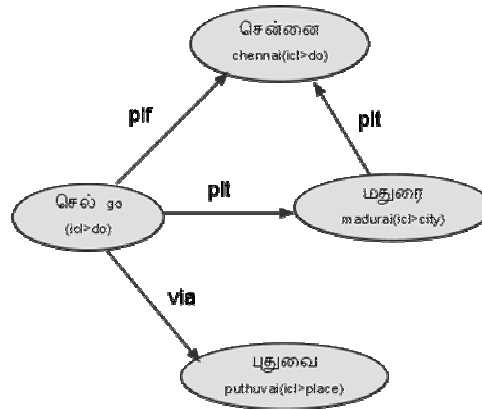
இந்த சொல் மற்றும் பொருள் விரிவாக்கம், தரவு மாற்றம் (data conversion) மற்றும் சுட்டு ஆக்க வரிசை (indexer) என்னும் இரு முக்கியமான செயல்முறைகள் மூலம் பெறப்படுகிறது. தரவு மாற்றம் என்பது 'இணையத் தவழி' மூலம் பெறப்பட்டு தமிழ்க் கோப்புகளை உ.இ.மொழி கோப்புகளாக மாற்றப்படுகிறது. அக்கோப்புகளில் உள்ள வாக்கியத்தை ஒவ்வொன்றாக எடுத்து வாக்கியத்தில் உள்ள சொல் மற்றும் சொல்லிற்கான உ.இ.மொ வின் சொல்லகராதி, சொற்பொருள், உ.இ.மொழியின் இலக்கண விதிகளுக்கு உட்படுத்தி இரு சொல்லிற்கு இடையில் உள்ள தொடர்பு நிலை, சொல்லின் வகை, சொல்லின் அடையாள எண், கோப்பின் பெயர், வாக்கியத்தின் எண் போன்ற தகவல்கள் பெறலாம். கோப்புகளில் உள்ள வாக்கியத்தில் இருந்து தகவல்கள் பெறும் முறை 'மொழிமரமாற்றம்' (enconversion)[5]ஆகும்.

ஒரு வாக்கியத்தில் உள்ள சொற்களுக்கு உலக இணைய மொழியின் 'உலகச் சொல் மற்றும் அச்சொற்களுக்கு இடையிலான தொடர்பு நிலையை கீழ் காணும் படம் எண் 3.1 ல் காணலாம்

"சென்னையிலிருந்து புதுவை வழியாக மதுரைக்கு செல்லலாம்."

சொல்	உ.இ.மொ வின் உலகச் சொல்
சென்னை	chennai(icl>place)
புதுவை	pudivai(icl>place)
மதுரை	madurai(icl>city)
செல்	go(icl>do)

சென்னை என்ற தமிழ் சொல்லுக்கு உ.இ.மொ வின் ஒலிபெயர்ப்பு (transliteration) chennai(icl>place). இதில் chennai என்னும் ஆங்கிலச்சொல் உ.இ.மொ தலை சொல்லாகும், icl>place என்பது உ.இ.மொ நிபந்தனையாகும். இத்தகவல்கள் உ.இ.மொ வின் அறிவுத்தளத்திலிருந்து பெறப்படுகிறது. உ.இ.மொ நிபந்தனை(constraint) ஒரு சொல்லின் சூழ்நிலையை குறிக்கும், அச்சொல்லின் சூழ்நிலைக்கேற்ப பொருள் வேறுபடும்.

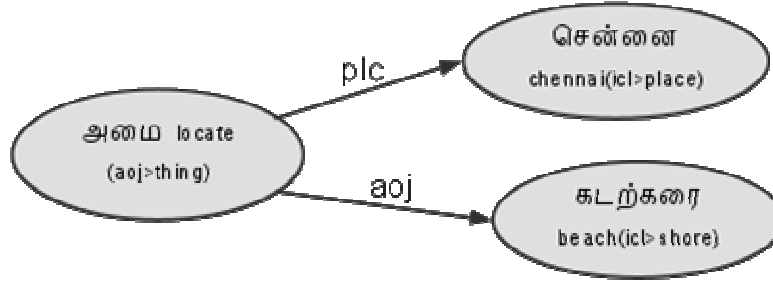


படம் எண் 3.1 சொற்களுக்கிடையிலான உ.இ.மொ வின் தொடர்பு நிலை வரை படம்

மேல்கண்ட படம் எண் 3.1ல் உ.இ.மொ வின் விதிப்படி செல் go(icl>do) <plf> சென்னை chennai(icl>place) என்பது 'சென்னையிலிருந்து செல்' என்ற பொருளை தரும். மதுரை madurai(icl>city) <plt> சென்னை chennai(icl>place) என்பது 'சென்னையிலிருந்து மதுரைக்கு' மற்றும் செல் go(icl>do) <plt> மதுரை madurai(icl>city) என்பது 'மதுரைக்கு செல்' என்ற பொருளை தரும். செல் go(icl>do) <via> புதுவை puduvai(icl>place) என்பது 'புதுவை வழியாக செல்' என்ற பொருளை தரும். இவ்வாறாக மேல் காணும் வாக்கியம் நான்கு தொடர்பு நிலைகளைக் கொண்டுள்ளது.

"சென்னையில் கடற்கரை அமைந்துள்ளது."

சொல்	உ.இ.மொ வின் உலகச் சொல்
சென்னை	chennai(icl>place)
கடற்கரை	beach(icl>shore)
அமை	locate(aoj>thing)



படம் எண் 3.2 சொற்களுக்கிடையிலான உ.இ.மொ வின் தொடர்பு நிலை வரை படம்

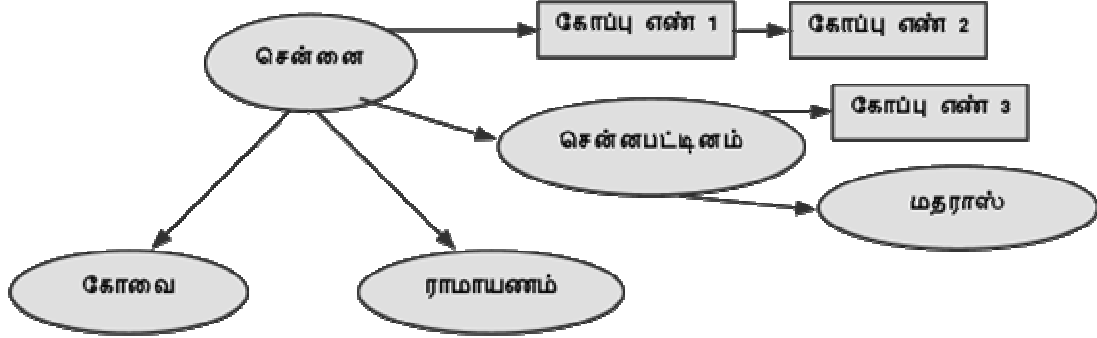
மேல்கண்ட படம் எண் 3.2ல் உ.இ.மொ வின் விதிப்படி அமை locate(icl>place), <plc> சென்னை chennai(icl>place) மற்றும் அமை locate(icl>place), <aoj> கடற்கரை beach(icl>shore) என்னும் இரு தொடர்பு நிலைகள் அமைந்துள்ளது.

3.1 தகவல் அமைப்பு

சொல் விரிவாக்கத்திற்கு இரும் தேடல் மர[4] (Binary Search Tree) தகவல் அமைப்பின் மூலம் விரிவாக்க சொற்கள் சேமிக்கப்படுகிறது. இரும் தேடல் மரத்தில் விரிவாக்கச் சொல் சுட்டுமுகவரி மூலம் சேமிக்கப்படுகிறது. சுட்டுமுகவரியை கொண்டு சேமிப்பதன் மூலமாக இரும் தேடல் மரத்தில் ஊர்ந்து சென்று மைய முனையில் உள்ள தகவல்களை எளிதில் பெற முடியும். இரும் மரத்தின் எல்லா முனையிலும் தொகுப்புப் பட்டியல்கள் இணைக்கப்பட்டிருக்கும். இரும் மரத்தின் முனையில் ஒரு சொல்லுக்கான உலக சொல்லின் சுட்டுமுகவரி (hash code) அமைந்திருக்கும். தொகுப்புப் பட்டியலில் சொல்லின் விரிவாக்கச்சொல், இரு சொற்களின் தொடர்பு நிலை, சொல் அமைந்திருக்கும் கோப்பு எண், வாக்கியத்தின் எண் போன்ற தகவல்கள் சேமிக்கப்பட்டிருக்கும்.

உதாரணத்திற்கு கீழ் காணும் படம் எண் 3.3ல் இரும் தேடல் மரத்தில் மைய முனையில் சென்னை, கோவை, ராமாயணம் போன்ற சொற்களின் சுட்டுமுகவரியும், அச்சொல்லின் விரிவாக்கச்சொல், இரு சொல்லின்

தொடர்பு நிலை, சொல் அமைந்திருக்கும் கோப்பு எண், வாக்கியத்தின் எண் போன்ற தகவல்கள் மைய முனையின் தொகுப்புப் பட்டியலில் சேமிக்கப்பட்டிருக்கும். மைய முனையில் தொடர்புடைய மற்றொரு முனையில், மைய முனையின் சொல்லுக்குரிய வேறு பொருள் கொண்ட சொல்லும் அதன் தொகுப்புப் பட்டியலில் அச்சொல்லுக்குரிய விரிவாக்கச்சொல்லும் அதன் தகவல்களும் சேமிக்கப்பட்டிருக்கும். இங்கு சென்னை என்னும் சொல் மைய முனையிலும், அதன் தொடர்புடைய மற்றொரு முனையில் சென்னை என்ற சொல்லுக்கு வேறு பொருள் கொண்ட சொற்கள் சென்னப்பட்டினம் மற்றும் மதராஸ் சேமிக்கப்பட்டிருக்கும்.



படம் எண் 3.3. இரும தேடல் மர தகவல் அமைப்பின் வரை படம்.

3.2 சுட்டு ஆக்க வரிசை

தரவு மாற்றத்தின் மூலமாக பெறப்படும் தகவல்களை, சொல்(சொ), சொல் மற்றும் அதன் தொடர்பு நிலை(சொ-தொ) , இரு சொல் மற்றும் இரு சொற்களுக்கு இடையே உள்ள தொடர்பு நிலை என்று(சொ-தொ-சொ) மூன்று பிரிவுகளாக பிரித்து சொ, சொ-தொ,சொ-தொ-சொ என்று மூன்று இரும மரங்களில் தகவல்களை சேமிக்கும் முறை சுட்டு ஆக்க வரிசை (index) ஆகும். சுட்டு ஆக்க வரிசையானது இரும மர தகவல் அமைப்பை கொண்டு உருவாக்கப்பட்டதாகும். இரும தேடல் (Binary Search) மூலமாக மிக எளிமையான முறையில் தகவல்களை பெறலாம்[2].

4 முடிவுகள்

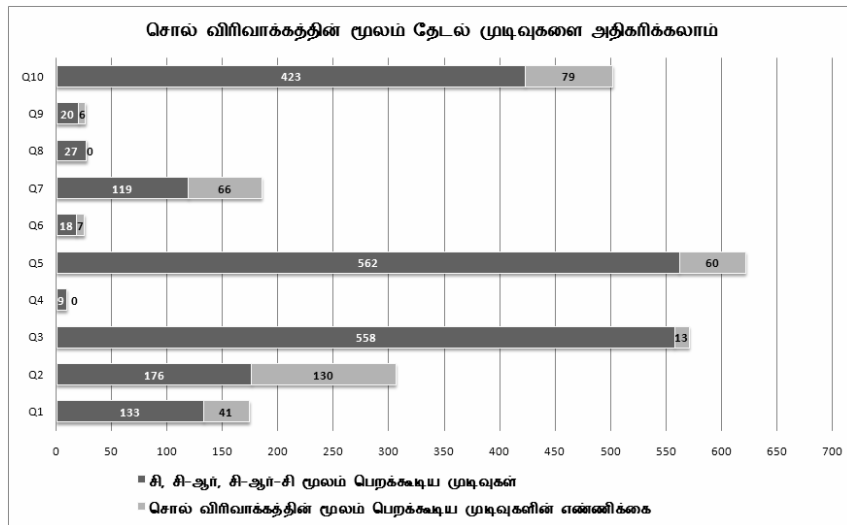
சொ-தொ-சொ இரும மரத்தில் உள்ள தகவல்களை கொண்டு சொல் விரிவாக்கம் செய்யப்படுகிறது. இரும மரத்தின் மைய முனையில் சொல்லுக்குரிய உலகச் சொல்லையும், அச்சொல்லின் பொருளுடைய வேறு சொல்லை, அதன் தொடர்புடைய மற்றொரு முனையில் சேமிப்பதன் மூலமாக சொல் மற்றும் சொல் சார்ந்த பொருள் விரிவாக்கங்களை. எளிதில் பெற முடியும். சொ-தொ-சொ வின் அதிர்வெண்ணை கொண்டு தர வரிசை அமைப்பதன் மூலம் ஒரு சொல்லுக்கு பொருத்தமான விரிவாக்கச் சொல்லை பெறமுடியும். இந்த புதிய முறையின் மூலம் விஷ்ணு என்னும் சொல்லுக்கு உ.இ.மொ வின் தொடர்பு நிலை விதியின் மூலம் விஷ்ணு என்னும் சொல்லுக்கும் விஷ்ணுவை ஒத்த பொருளுடைய வேறு சொற்களான கிருஷ்ணர், வெங்கடாசலபதி போன்ற சொற்களுக்கு சொல் விரிவாக்கம் செய்யப்பட்டுள்ளதை கீழ் காணும் அட்டவணை எண் 3.2ல் காணலாம். இந்த சொல் விரிவாக்கத்தில் pos மற்றும் and என்பது உ.இ.மொ வின் தொடர்பு நிலையாகும். "அக்லோமேரடிவே தொகுதி வழிமுறையின்" மூலம் விஷ்ணு என்னும் சொல்லுக்கு மட்டுமே சொல் விரிவாக்கம் செய்யமுடியும். அதன் விரிவாக்கசொல்லை கீழ் காணும் அட்டவணையில் காணலாம்.

சுட்டு ஆக்க வரிசை பொருள் விரிவாக்கம்	அக்லோமேரடிவ்
விஷ்ணு <pos> அவதாரம்	விஷ்ணு - இந்தியா
விஷ்ணு <pos> ஆலயம்	விஷ்ணு - இந்து
விஷ்ணு <pos> கோயில்	விஷ்ணு - கடவுள்
விஷ்ணு <pos> கோவில்	விஷ்ணு - காலம்
விஷ்ணு <pos> வழிபடு	விஷ்ணு - கோயில்
விஷ்ணு <and> பிரம்மா	விஷ்ணு - சிவன்
கிருஷ்ணர் <pos> அவதாரம்	
வெங்கடாசலபதி <pos> ஆலயம்	

படம் எண் 3.2. பொருள் விரிவாக்கம் மற்றும் அக்லோமேரடிவே தொகுதி வழிமுறையின் சொல் விரிவாக்கத்தின் முடிவு அட்டவணை.

4.1 செயலாற்றல்

'பொருள் தேடலுக்கான' விரிவாக்கச் சொல் உருவாக்க முறையை முப்பத்து மூன்று ஆயிரத்திற்கும் மேற்பட்ட சுற்றுலா சம்மந்தமான கோப்புகள் மூலம் பகுப்பாய்வு செய்யப்பட்டது .பகுப்பாய்வின் முடிவில் ஒரு சொல்லுக்கு, கோரியின் 'பொருள் விரிவாக்கம்' முறையின் மூலம் கூடுதலான கோப்புகளின் வளர்ச்சியை கீழ் உள்ள வரை படம் எண் 4.1ல் காணலாம். இரும் தேடல் மரத்தில் சேமிப்பதன் மூலம் விரிவாக்கச் சொல்லை மிக எளிமையாகவும், விரைவாகவும் பெற முடியும். கோப்புகளின் எண்ணிக்கை அதிகரிக்கும் போது சொல் மற்றும் பொருள் விரிவாக்க சொல்லும் அதிகரிக்க வேண்டும். இப்புதிய முறையை சோதிக்க "இணைய தவழி" மூலம் பெறப்பட்ட கோப்புகளை சோதனை செய்தோம். முதலில் 33,721 கோப்புகளுக்கு 1,36,913 விரிவாக்க சொல் பெறப்பட்டது. பின்னர் 41,721 கோப்புகளுக்கு 1,88,618 விரிவாக்க சொல் பெறப்பட்டது.

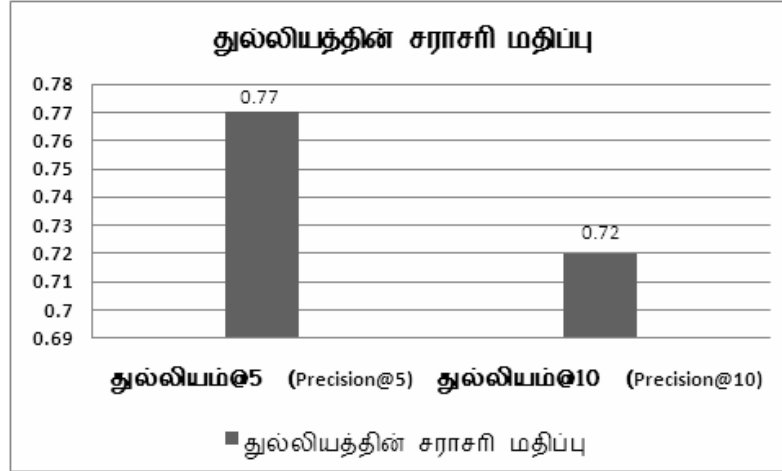


படம் எண் 4.1. பொருள் விரிவாக்கத்தால் அதிகரிக்கும் முடிவுகளின் எண்ணிக்கை

4.2 துல்லியம்

துல்லியம் என்பது இணையதேடல் முடிவுகளில் பொருத்தமான முடிவுகளை பெற உதவும் அளவுகோல் ஆகும். கீழ் காணும் சமன்பாட்டை பயன்படுத்தி கணக்காய்வு செய்ய வேண்டும்[12]. கோரி தேடலில் சொல் விரிவாக்கம் துல்லியத்தை கீழ் காணும் படம் எண் 4.2 ல் காணலாம். இதில் துல்லியம்@5 என்பது கோரி இணையத்தேடலில் முதல் 5 முடிவுக்கான துல்லியத்தின் சராசரி. துல்லியம்@10 என்பது கோரி இணையத்தேடலில் முதல் 10 முடிவுக்கான துல்லியத்தின் சராசரி

$$\text{துல்லியம்} = \frac{\text{பொருத்தமான கோப்புகள்} \cap \text{பெறப்பட்டக் கோப்புகள்}}{\text{பெறப்பட்டக் கோப்புகள்}}$$



படம் எண் 4.2. சொல் விரிவாக்க முடிவுகளின் துல்லிய சராசரி மதிப்பு.

5. முடிவுரை

ஒரு தகவல் தேடலின் செயல்திறன், தேடல் முடிவுகள் கொண்டு நிர்ணயம் செய்யப்படுகிறது. சொல் சார்ந்த தேடல்களில், சொல்லுக்கான கோப்பு இல்லாவிடில் சொல்லின் பொருளுடைய கோப்பு அல்லது அச்சொல்லின் பொருளுடைய வேறு சொற்கள் சார்ந்த கோப்புகள் பெறுவது இயலாத நிலையாக உள்ளது. இது தமிழ் தகவல் தேடலில் ஒரு குறைபாடாக இருந்து வருகின்றது. இந்த குறைபாட்டை நீக்க கோரியின் சொல் மற்றும் பொருள் விரிவாக்க முறையின் மூலம் நிவர்த்தி செய்யமுடியும். புதிய சொல் மற்றும் பொருள் விரிவாக்க முறையின் மூலம் இணைய தேடல் முடிவுகளின் துல்லியம் அதிகரிக்கும். வருங்காலத்தில் பொருள் விரிவாக்கத்தின் துல்லியத்தை மேலும் அதிகரிக்கவும், இம்முறையை பயன்படுத்தி தமிழ் மொழி மட்டுமின்றி பல மொழிகளில் உள்ள தகவல்களை பெறுவதற்கு வழிவகை செய்யப்படும்.

Reference

1. UNDL. 2009. Universal Networking Digital Language. <http://www.undl.org>, last accessed date 12 March 2010.
2. Subalalitha, T.V, G., Parthasarathi, R., and Karky, M. 2008. corex: a concept based semantic indexing technique. swm-08.

3. Cluster Analysis. http://en.wikipedia.org/wiki/cluster_analysis, last accessed date 12 March 2010.
4. Weiss, M. A. February 2006. Data Structures and Algorithm Analysis in C++. number ISBN-13:9780321441461. addison wesley.
5. T.Dhanabalan, K.Saravanan, and T.V.Geetha. 2002. Tamil to UNL Enconverter, ICUKL, Goa, India.
6. Agissilaos Andreou, 2005. Ontologies and Query Expansion. Master of Science, University of Edinburgh.
7. Pattabhi R.K Rao and Sobha.I, FIRE 2010. Cross Lingual Information Retrieval Track: Tamil-English.
8. Reginald Ferber,1997. Automated Indexing with Thesaurus Descriptors: A Cooccurrence Based Approach to Multilingual Retrieval.
9. WordNet. <http://wordnet.princeton.edu>, last accessed date 12 March 2010.
10. WordNet. <http://en.wikipedia.org/wiki/wordnet>, last accessed date 12 March 2010.
11. Agglomerative Clustering. <http://www-servers.cs.umn.edu/~sushrut/research/pub/cover/node23.html>. last accessed date 12 March 2010.
12. Information Retrieval Performance Measure. http://en.wikipedia.org/wiki/information_retrieval, last accessed date 12/04/2010.

CoRe - A Framework for Concept Relation based Advanced Search Engine

T V Geetha, Ranjani Parthasarathi & Madhan Karky
{tv_g@hotmail.com, rp@annauniv.edu, madhankarky@gmail.com}
Department of Computer Science & Engineering
College of Engineering Guindy
Anna University

Abstract

The number of Tamil documents is growing rapidly every second. Blogs, news portals, infonets and socialnets have considerably contributed to this growth. Traditional keyword based searches such as Google or Bing, primarily developed for English are now being used to search Tamil documents. These search engines do not have a full-fledged support for Tamil language. We present CoRe, the world's first framework for a concept relation based search engine for Tamil. CoRe search, unlike traditional keyword based searches, identifies the concepts in a document and their relationship with each other using UNL (Universal Networking Language). UNL is essentially a semantic relation based intermediate representation that has been used in this work for search index representation. This paper presents the primary components of the CoRe framework, namely, EnCoRe(enconverter), CoReX(indexer) and CoReS(Search & Rank) modules. The CoRe based search is tested with over 50,000 crawled web documents and the results are compared against a traditional keyword based search algorithm over the same set of documents. It is observed that the UNL based search has a precision accuracy of 75%

8. Introduction

With UNICODE being accepted internationally as a standard for representing text on web sites, search engines are now indexing the pages independent of the language. Popular search engines like Google, Bing and Yahoo now readily index Tamil documents. Only Google manages to include certain language dependent features such as stemmer for Tamil. Even though keyword based search seems very convenient system to retrieve web documents it fails in two major issues. The first one is in understanding the document. The second one is in understanding what the user wants.

Let us start with a very simple example where a user wants to know about a temple in the city of Madurai. The user who does not know the actual name of the temple will use keywords Madurai and Temple in a search engine, which is very likely to retrieve documents that have mentions of both these keywords. What the traditional search engine will not be able to retrieve are documents that talk about Meenakshi temple which do not contain any of the search keywords. This failure of traditional search engine expects the user to reach these documents in their second or third search attempts with different keywords.

Personalised search engines aim to solve this problem from the user side. Concept based search engines aim to solve the same problem from both the ends. In this paper we propose CoRe, a framework for a Concept - Relation Based Search. The CoRe framework aims at understanding the web documents by

indexing them based on concepts and their relations rather than indexing the keywords and their frequencies. The framework was implemented and tested with over 50,000 documents from tourism domain. The search results were compared against traditional keyword based search for precision and relevance.

This paper is organized into five sections. The second section provides a brief survey of literature relevant to concept based search subsystems. The third section presents the CoRe framework and its components. The fourth section discusses the results of CoRe search engine and compares it against traditional keyword based search for precision and relevance. The fifth section summarises the paper, presents the work in progress and discusses future directions.

2. Background

Meaning based or concept based search and indexing techniques have been proposed for English and other languages. The main purpose of such indexing techniques is cross lingual information retrieval. Universal Networking Language (UNL), proposed in our CoRe framework for conversion of Tamil documents, is described in [3]. UNL is an interlingua framework that was originally designed to aid the machine translation process[3]. An UNL converter interprets natural language and converts it into an intermediate representation that utilizes language independent semantic concepts and relations prescribed by UNL. The conversion from the source language to UNL analyses the source language and utilizes language specific linguistic rules for building the UNL representation. On the other hand the UNL deconverter transforms UNL representation to a target language. For this purpose the deconverter uses language specific generation rules to produce the target language.

Kang in [4] proposes a indexing technique where he indexes documents based on the concepts identified in the document. A similar work where medical images are conceptually indexed can be found in [5]. Chau and Yeh in [2] proposes another conceptual index again based on concepts. Here Chau and Yeh have designed the index especially for cross lingual text retrieval. Surve et al., in [7] propose AgroExplorer, a meaning based multi lingual search system. Though these systems have the advantage of concept based indexing they do not investigate deep into the relations between the concepts in a sentence.

3. CoRe Search Framework

CoRe search framework presented in figure 1, can be divided into two major divisions, online and offline, in terms of the time of processing. Three major subsystems EnCoRe, CoReX and CoReS form the backbone of the framework providing the major functionality. Tamil language tools are separated from rest of the system to offer language dependent services such as analysers and generators. This section describes the various components of the framework in detail.

3.1 Offline Processing

The offline mode comprises of operations related to crawling Tamil web documents relevant to a particular domain, processing the raw document for extracting sentence constituents, converting the sentence constituents to corresponding UNL graphs and building the concept relation index. This is a periodic scheduled process and will be modifying the index incrementally as new documents are being crawled.

3.2 Online Processing

A user's query is processed, expanded and converted to UNL graph(s) and sent to a search and ranking subsystem where the documents that match concept relation similarity are ranked and sent for output processing. The output processing module formats the output, generates snippets and summary for the retrieved documents and sends it to the user.

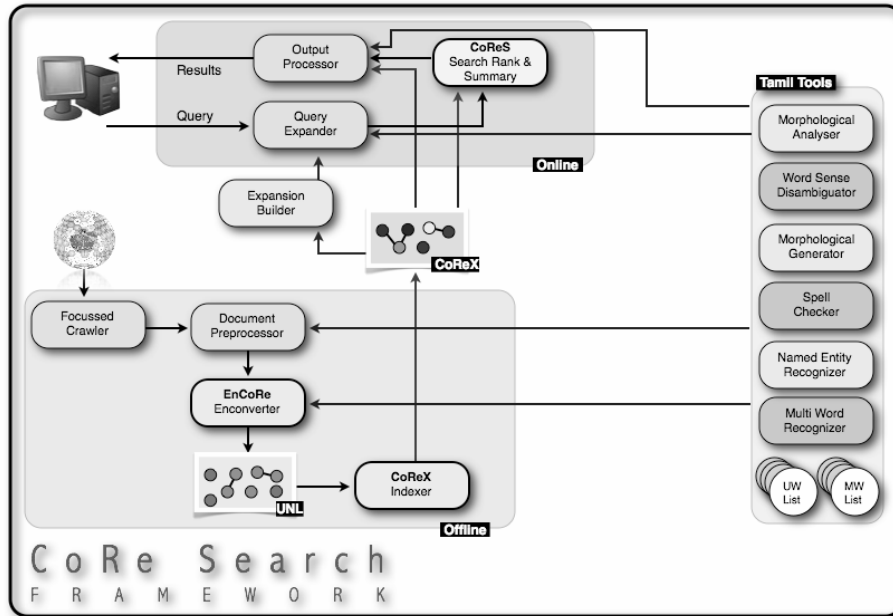


Fig 1 : CoRe Search Framework

3.3 Tamil Tools

The Tamil tools package offers a set of language processing services for different components of the CoRe framework. Morphological Analyser [1] is used by most of the components for morphologically recognize a word. The rules for the EnCoRe subsystem depends on the results of the morphological analysis. Word Sense Disambiguator(WSD), a rule based tool, is used to resolve the meaning of a word based on it's context. This tool is used both by EnCoRe and Query Expander. Morphological Generator is used for generating natural language sentences. Output Processor uses Morphological Generator to generate natural language summary for a given document. Query Expander and EnCoRe use spell checker to auto correct typos and basic morphological errors. Named Entity Recogniser and Multiword Recogniser are used to identify named and multiword entities respectively. These rule based tools are used by EnCoRe and Query Expander to identify entities. The Tamil Tools also comprises of a Universal Word list(UW list) [3] and a Multiword List(MW list). These lists carry the domain related UW and MW words respectively.

3.4 Focussed Crawler

Tamil web documents specific to a certain domain are fetched into the system by a Focussed crawler. The domain specific words are fed to the crawler along with a seed URL list. The documents fetched by the crawler are sent to the Document Pre-processor for further processing.

3.5 Document Pre-processor

Document Pre-processor parses Tamil documents in HTML format for textual content removing links and other unwanted tags. The Pre-processor also identifies important sentence constituents and send the sentence constituent along with a document id to the EnCoRe subsystem.

3.6 EnCoRe

The EnCoRe subsystem forms the heart of the CoRe framework. Here a sentence constituent is passed to a rule based system to identify the various concepts in the constituent and the rules are used to identify one of the 44 UNL relations [3]. EnCoRe uses language processing tools such as the Morphological Analyser to recognize various morphological suffices of a word and uses this information along with syntax and semantics to identify the relationship between concepts. UNL graphs are generated for every sentence constituent. The UNL graph is then sent to CoReX indexer along with information such as document ID, positional index and original keyword, its frequency in the document etc. to be used by the CoReS subsystem.

3.7 CoReX Indexer

The CoReX Indexer subsystem presented in [6] stores and manages the UNL graphs in three different indices. Concept only index(C index), Concept-Relation index(CR index) and Concept-Relation-Concept index(CRC index) are the three indices maintained by the CoReX indexer. The UNL graphs are stored in the indices by their concept ids for efficient retrieval. The CoReX index structure and efficiency analysis is provided in [6].

3.8 Query Expander & Expansion Builder

Any user query is directly sent to the Query Expander module that expands the query using the data from Expansion Builder. The Expansion Builder uses the CoReX index to build on-the-fly similarity thesaurus and co-occurrence list. Query Expander enconverts the user query to a UNL graph for the CoReS subsystem, using the information provided by the Expansion builder.

3.9 CoReS

CoReS subsystem provides the CoRe framework with the functionality of searching and ranking documents based on concepts and relations. Unlike traditional algorithms that scores pages based on terms and their frequencies, CoReS subsystem ranks a document based on the concepts and the type of relations that exist between those concepts.

3.10 Output Processor

Results from CoReS subsystem would be a set of documents and their corresponding weight with respect to the user query and expansions. Output processor formats the result page by generating snippets highlighting the identified concepts and also generates a template-based summary for every page. For instance in a tourism domain, the page summary would contain information about tourist spots, contact

numbers, animals, hotels, transport and more. Similarly a summary for a health domain may contain hospital info, medicine information, emergency contacts, symptoms of a disease and so on.

4. CoRe Search Results & Analysis

A search engine based on CoRe framework was implemented by modifying the Nutch open source search engine. Over 50,000 documents in tourism domain were crawled and converted and indexed for search. An implementation of a Tamil keyword based search built with Nutch over the same set of documents and domain is taken for comparison.

4.1 Matrix Layout



Fig 2 : Matrix Layout

In this paper, we propose a Matrix Layout for displaying the results to the user. The Matrix Layout shown in figure 2, displays the results in a 3X2 matrix cells with each cell corresponding to a class of results based on the concepts and relationship between concepts. Figure 2 displays the results for the query தஞ்சை பிரகதீஸ்வரர் கோயில். (thanjai birakatheeswarar koayil) The first cell displays the results pertaining to the concepts that contain the actual keywords and sorted by the relation they have between them. Second cell identifies results that contain at least one concept with the actual keyword. The third cell identifies documents that will not be identified by traditional keyword search where none of the concepts contain keywords. In this case it retrieves documents with தஞ்சாவூர் & கோவில்(thanjaavoor & koavil) both of which do not form part of the actual query. The fourth and fifth cells are based on expansions of the query. Here they display results corresponding to பெரிய கோவில் & திருக்கோவில் (periya koavil & thirukkoavil). The final cell identifies the place associated with the query term to display the map of the corresponding place. The snapshot presented in figure 2 is from our கோரி(coree) search engine implemented from the CoRe framework.

4.2 Performance Evaluation

Precision of documents can be computed using the formula given below[8]. We compute the precision of documents for the first 5, first 10 and first 20 documents.

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

The average precision and mean average precision for a set of queries will indicate the performance of the system.

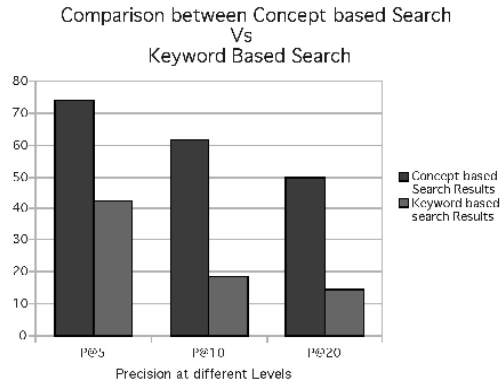


Fig 3 : Average Precision Comparison

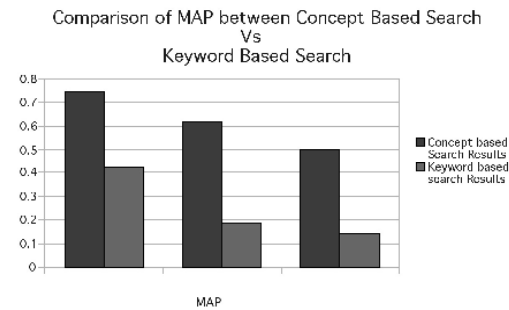


Fig 4 : Mean Average Precision Comparison

For a set of 100 queries, precision is calculated at three levels for both the CoRe based search and keyword based search. The results are presented as a graph in figure 3. The mean average precision (MAP) [8] score comparison of the two search paradigms is given in figure 4.

8. Conclusion and Future Work

This paper describes CoRe, a framework for concept relation based search in Tamil. Different subsystems and components of the framework are described in detail. Results from an implementation of CoRe framework is provided and is compared against traditional keyword based search results. The rules of EnCoRe subsystem are very much domain specific. Expansion of the search engine to adapt to other domains will be future work. Integrating cross lingual results and providing cross lingual summary for the results will take this work to its next level.

References

1. Anandan, R. Parthasarathi, and Geetha, *Morphological Analyser for Tamil*. ICON 2002, 2002.
2. Chau, R. and C.-H. Yeh, *Fuzzy Conceptual Indexing for Concept-Based Cross-Lingual Text Retrieval*. IEEE Computer Society, 2004. 8(5).
3. Foundation, U., *The Universal Networking Language (UNL) Specifications Version 3 3ed*. December 2004: UNL Cente UNDL Foundation
4. Kang, B.-Y. *A Novel Approach to Semantic Indexing Based on Concept*. In *41st Annual Meeting on Association for Computational Linguistics* 2003. Japan.
5. Lacoste, C., et al., *Inter-Media Concept-Based Medical Image Indexing and Retrieval with UMLS at IPAL*. Evaluation of Multilingual and Multi-modal Information Retrieval,, 2007. 4730: p. 694-701.

6. Subalalitha, et al. *CoReX : A Concept Based Semantic Indexing Technique*. In *SWM-08*. 2008. India.
7. Surve, M., et al. *AgroExplorer: a Meaning Based Multilingual Search Engine*. In *International Conference on Digital Libraries*. 2004. Delhi, India: Sannella.
8. Andrew, T. and S. Falk. *User performance versus precision measures for simple search tasks*. In *29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* 2006. Seattle, Washington, USA.

15

கையடக்கக் கருவிகளில் தமிழ்



தமிழில் செல்லிட ஆளுகை

அண்ணாகண்ணன்

(annakannan@gmail.com)

முனைவர் பட்ட ஆய்வாளர், தமிழ்த் துறை, பச்சையப்பன் கல்லூரி

இந்தியத் தகவல் தொடர்புத் துறையில் செல்பேசிகள், புரட்சிகரமான மாற்றங்களை உருவாக்கி வருகின்றன .இந்தியாவில் மார்ச்சு 2010 வரை 58. 43கோடி செல்பேசிச் சந்தாதாரர் உள்ளனர்.¹ ஒருவரே ஒன்றுக்குமேல் செல்பேசி கள் வைத்துள்ளதால் ,மாநகரங்களில் %100க்கு மேல் செல்பேசிகள் உள்ளன.

கைக்கு அடக்கமாய் இருக்கும் செல்பேசிகள், உலகையே நம் கைக்குள் அடக்கும் வல்லமை படைத்தவை . இதன் மூலம், இருந்த இடத்திலிருந்தே உலகை ஆள முடிகிறது .ஒரே செல்பேசி, நூற்றுக்கணக்கான வசதிகளை அளிக்கிறது .குறுங்கணினியாக மாறியதோடு ,புவிநிலை உணர் திறனுடன் தொலைநிலை இயக்கியாகவும் செயல்புரிகிறது .ஒரே செல்லில் 3 ,2, 4, 5 சிம் அட்டைகள் பொருத்தும் வசதியும் வந்துள்ளது . இரண்டாம் தலைமுறை, மூன்றாம் தலைமுறை, நான்காம் தலைமுறை செல்பேசிகளும் ஐபோன், ஐபாட் உள்ளிட்ட கருவிகளும் பரவி வருகின்றன. வாமனர் பேருருக் கொண்டது போல் ,செல்பேசியின் திறன் வளர்ந்துகொண்டே செல்கிறது.

அதே நேரத்தில் அதன் விலை, குறைந்து வருகிறது 500 .ரூபாய்க்கே புதிய செல்பேசிகள் கிடைக்கும் நிலை உருவாகியுள்ளது.² இதனால் அனைத்துத் தரப்பு மக்களும் இவற்றைப் பயன்படுத்தி வருகின்றனர் .அதிகப் பயன்பாடு காரணமாக செல்பேசி கோபுர எண்ணிக்கை, பெருகி வருகிறது. சென்னையில் 2,500 தனியார் செல்பேசிக் கோபுரங்கள் இருப்பதாகக் கணக்கிடப்பட்டுள்ளது.³

செல்பேசி எண்ணை மாற்றாமல், செல்பேசிச் சேவையாளரை மட்டும் மாற்றிக்கொள்ளும் வசதி, 2010 மே மாதத்தில் அறிமுகமாக உள்ளது.⁴ இதனால் சேவைகளின் தரம் மேலும் அதிகரிக்கும். செல்பேசிகள் , மக்களிடையே மகத்தான செல்வாக்கைப் பெற்றுவிட்டன .பலரும் அதை விட்டுச் சற்றும் விலகுவதில்லை . இதனால் ,அரசுத் துறைகள் உள்பட அனைத்துத் துறைகளும் செல்பேசி வாயிலாகச் சேவைகளை அளிக்கத் தொடங்கிவிட்டன.

செல்பேசிக்கு வரும் மின்னஞ்சல்

இந்தியாவில் அகண்ட அலைவரிசை இணைப்புகளை விட 60 ,மடங்குகளுக்கும் அதிகமாகச் செல்பேசி இணைப்புகள் உள்ளன .ஐ.பி.ஆர்.எஸ். வசதி இல்லா நிலையிலும் இந்தச் செல்பேசி வாடிக்கையாளர் அனைவரும் இப்பொழுது ,இணையத்தை அணுக இயலும் .இதன்படி, மின்னஞ்சலானது, குறுஞ்செய்தியாக உரியவரின் செல்பேசிக்கு அனுப்பப்படும் .பெரிய மின்னஞ்சலாக இருப்பின், பல குறுஞ்செய்திகளாக அது வந்து சேரும் .இந்தச் சேவையை <http://m3m.in> என்ற நிறுவனம், வழங்குகிறது .எனவே முதல் முறை ஒரு மின்னஞ்சல் முகவரியை உருவாக்கிவிட்டால் போதும் .அதன் பின்னர், செல்பேசியைத் தேடி மின்னஞ்சல்கள் வந்து சேரும் .இதன் மூலம் மறைமுகமாக இணையப் பயன்பாடு வளரும் வாய்ப்புண்டு . மின் சமூக இடைவெளியைக் குறைக்கவும் செல்பேசிகள் பெரிதும் உதவும்.

1. செல்லிட ஆளுகை - ஓர் அறிமுகம்

அரசுச் சேவைகள் ,தகவல்கள் ஆகியவற்றை அனைத்து வகைக் கம்பியில்லாத் தொடர்பு நுட்பங்கள் வாயிலாக வழங்குவதையும் பெறுவதையும் செல்லிட ஆளுகை எனச் சுருக்கமாகக் கூறலாம் .

மின்னாளுகையின் தொடர்ச்சியாக இஃது உருவானாலும், செல்லிட ஆளுகை தனித்து இயங்கவும் வல்லது . நடப்பு நிலையில் மின்னாளுகைச் சேவைகளைச் செல்பேசி உள்ளிட்ட கருவிகள் வாயிலாக வழங்குவது செல்லிட ஆளுகை எனப்படுகிறது.⁵

இந்தியாவில் செல்லிட ஆளுகை

இந்தியாவில் மத்திய -மாநில அரசுத் துறைகள், மாநகராட்சிகள், நகராட்சிகள், மாவட்ட நிர்வாகம், மின் வாரியம், வங்கிகள், தொலைபேசி நிறுவனங்கள், கல்வித் துறை உள்ளிட்ட பலவும் மின்னணு வசதிகளைப் பயன்படுத்தத் தொடங்கிவிட்டன. இவை, கட்டண நினைவூட்டல், கட்டணத்தைப் பெற்றுக்கொண்ட விவரம், புகார், அதன் மீதான நடவடிக்கை, புதிய சேவைகள், சலுகைகள் ...எனப் பலவற்றுக்கும் குறுஞ்செய்திகளை அனுப்பி வருகின்றன.

அரசு அலுவலகங்களில் கால்கடுக்க நீண்ட நேரம் வரிசையில் நின்று, பெற்ற விவரங்களை விரல் சொடுக்கில் பெற முடிகிறது .உணவு இடைவேளை, தேநீர் இடைவேளை, சனி, ஞாயிறு, விடுமுறை நாட்கள் ஆகிய எல்லாக் காலங்களிலும் அரசினைத் தொடர்புகொள்ளக்கூடிய வாய்ப்பினை மக்கள் பெற்றுள்ளனர் . பேருந்திலும் தொடர்வண்டியிலும் செல்லும்பொழுதே இவ்வாறான சேவைகளைப் பெற முடிகிறது . செல்பேசி வழியே கட்டணம் செலுத்தும் வசதியும், பல துறைகளில் அறிமுகப்படுத்தப்பட்டுள்ளது.⁶

2. தமிழகத்தில் செல்லிட ஆளுகைச் சேவைகள்

தமிழகத்தில் மத்திய -மாநில அரசுத் துறைகள் ,பல்வகையான செல்லிட ஆளுகைச் சேவைகளை அளித்து வருகின்றன .அவற்றுள் சில இங்கே:

செல்பேசியில் தேர்வு முடிவுகள்

12 ,10ஆம் வகுப்புகளின் தேர்வு முடிவுகளையும் மதிப்பெண் பட்டியலையும் செல்பேசி வழியே பெறும் முறை, தமிழகத்தில் நடைமுறைக்கு வந்துள்ளது .அரசிடமிருந்து தேர்வு முடிவுகளைப் பெற்றுப் பல்வேறு இணையத்தளங்களும் வெளியிட்டு வருகின்றன.⁷ இதற்குச் சிறிதளவு கட்டணம் உண்டு.

தொடர்வண்டித் துறையில் குறுஞ்செய்திச் சேவை

139என்ற எண்ணுக்குக் குறுஞ்செய்தி அனுப்பி ,பயணச் சீட்டு இருப்பு நிலை ,கட்டண விவரம் , காத்திருப்புப் பட்டியலின் நிலை ,குறிப்பிட்ட தொடர்வண்டி தற்போது எந்த இடத்தில் உள்ளது ஆகியவற்றை அறியலாம் .இரண்டு குறிப்பிட்ட நிலையங்களுக்கு இடையே இயங்கும் தொடர்வண்டிகள் , அவற்றின் கால அட்டவணை உள்ளிட்ட விவரங்களையும் அறியலாம்.

சென்னை மாநகராட்சிக்குப் புகார் அனுப்ப

சென்னை மாநகராட்சிக்குப் புகார், ஆலோசனை அளிக்க, 9789951111என்ற எண்ணுக்குக் குறுஞ்செய்தி அனுப்பலாம். மேலும் ,சொத்து வரி உள்ளிட்டவற்றையும் செல்பேசி வழியாகவே செலுத்தலாம்.⁸

புதிய குடும்ப அட்டை குறித்து, குறுஞ்செய்தியில் தகவல்

தமிழ்நாட்டில் புதிய குடும்ப அட்டை கேட்டு விண்ணப்பிப்பவர்கள், செல்பேசி எண், மின்னஞ்சல் முகவரி ஆகியவற்றைத் தெரிவித்தால் குடும்ப அட்டை தயாரானதும் குறுஞ்செய்தி, மின்னஞ்சல் வாயிலாகத் தகவல் தெரிவிக்கும் திட்டம், 2010ஜனவரி 1ஆம் தேதி முதல் நடைமுறைக்கு வந்துள்ளது.⁹

ஓய்வூதியதாரருக்குக் குறுஞ்செய்திச் சேவை

கணக்காயர் அலுவலகத்தில் ஓய்வூதியதாரரின் விண்ணப்பம் பெறும்போது ,ஓய்வூதிய வழங்காணை அனுப்பும்போது, விண்ணப்பத்தில் ஏதாவது குறையிருந்து, திருப்பி அனுப்பும்போது ஓய்வூதியதாரருக்கு

அது குறித்துக் குறுஞ்செய்தி அனுப்பப்படுகிறது .இந்தத் திட்டம் பொதுச் சேமநல நிதி கணக்கு முடிப்பு உள்ளிட்ட இதர பிரிவுகளுக்கும் விரிவுபடுத்தப்பட உள்ளது.¹⁰

போக்குவரத்துத் துறையில் குறுஞ்செய்திச் சேவை

போக்குவரத்துத் துறை தொடர்பான புகார்களையும் போக்குவரத்து விதிமீறல்களையும் விபத்துப் பற்றிய செய்திகளையும் குறுஞ்செய்தி வழியாகத் தெரிவிக்க முடியும் .இதற்காகச் சென்னை போக்குவரத்துத் துறை இணை ஆணையர் சுனில் குமார், 08123-98418என்ற எண்ணை 2006ஜூலை 20அன்று அறிவித்துள்ளார் . 100, 103, 42042300, 28521323ஆகிய முந்தைய எண்களும் பயன்பாட்டில் உள்ளன .'இந்த எண்ணை அழைத்துப் பேச வேண்டாம்; அதற்குப் பதில், குறுஞ்செய்திகளாக அனுப்புங்கள்' என சுனில் குமார் வேண்டியுள்ளார்.¹¹ மேலும் ,போக்குவரத்து நெரிசல் குறித்த தகவல்களைச் செல்பேசி சேவை வழங்கும் நிறுவனங்களின் உதவியுடன் குறுஞ்செய்திகளாக மக்களுக்குத் தெரிவிக்கத் திட்டமுள்ளது எனப் போக்குவரத்துத் துறையின் கூடுதல் ஆணையர் ஷகீல் அக்தர் தெரிவித்துள்ளார்.¹²

சுற்றுலாத் துறையில் குறுஞ்செய்திச் சேவை

தமிழ்நாடு சுற்றுலா வளர்ச்சிக் கழகத்தின் தங்கும் விடுதிகளைக் குறுஞ்செய்தி மூலம் முன்பதியலாம் .இந்த வசதியின் மூலம் பல்வேறு சுற்றுலாப் பயணத் திட்டங்களையும் முன்பதியலாம் .இதனைத் தமிழகச் சுற்றுலாத் துறை அமைச்சர் சுரேஷ்ராஜன் அறிவித்துள்ளார்.¹³

காவல் நிலையத்திற்குக் குறுஞ்செய்தியில் புகார்

காவல் நிலையத்திற்குக் குறுஞ்செய்தி மூலம் புகார் தெரிவிக்கும் திட்டம், புதுவையில் தொடங்கப்பட்டுள்ளது .புதுவையின் 4இல் ஒரு பகுதியில் இத்திட்டம் பரிசோதனை நிலையில் செயல்படுத்தப்படுகிறது .இதனைப் புதுச்சேரி உள்ளதுறை அமைச்சர் வலசராஜ், 21.02.2010 அன்று தெரிவித்துள்ளார்.¹⁴

பள்ளிக்கு மாணவர் வராவிட்டால், பெற்றோருக்குக் குறுஞ்செய்தி

பள்ளிக்கு மாணவர் வராவிடில், அது குறித்துப் பெற்றோரின் செல்பேசிக்குக் குறுஞ்செய்தி அனுப்பும் முயற்சியில் சென்னை மாநகராட்சி ஈடுபட்டுள்ளது .இது, நங்கம்பாக்கத்தில் உள்ள மாநகராட்சி பெண்கள் மேல்நிலைப்பள்ளியில் 9, 10, 11, 12 ஆகிய வகுப்புகளில் பயிலும் மாணவிகளிடம் நடைமுறைப்படுத்தப்பட உள்ளது .இந்தச் சோதனை முடிவுகளை அடுத்து, மாநகராட்சி நடத்தும் அனைத்துப் பள்ளிகளிலும் நடைமுறைப்படுத்தப்பட உள்ளது.¹⁵

திருவள்ளூர் /கன்னியாகுமரி மாவட்டங்களில் செல்பேசிவழிச் சேவைகள்

திருவள்ளூர் ,கன்னியாகுமரி மாவட்டங்களில் 54373என்ற எண்ணுக்கு குறுஞ்செய்தி அனுப்பி ,பிறப்புச் சான்றிதழ், இறப்புச் சான்றிதழ், இருப்பிடச் சான்றிதழ், சாதிச் சான்றிதழ், வருமானச் சான்றிதழ், நிலப்பட்டா விவரங்கள் ஆகியவற்றைப் பெறலாம் .புகார் மனுக்களையும் பதியலாம்.¹⁶ படச் செய்தி மூலம் புகார், கோரிக்கை அளிக்கவும் வாய்ப்புண்டு .

பிறப்புச் சான்றிதழ் பெற விரும்புபவர் TLR BC Date_of_Birth Baby_Name Place_of_Birth Parents_Name_and_Address என்ற வடிவில் விவரங்களை அனுப்ப வேண்டும் .இத்தகு குறுஞ்செய்திகளில் குறிப்பிடப்படும் குறுக்கங்களுக்கான விரிவு வருமாறு:

TLR - திருவள்ளூர் ,BC - பிறப்புச் சான்றிதழ் ,DC - இறப்புச் சான்றிதழ் ,IC - வருவாய்ச் சான்றிதழ் , RC - இருப்பிடச் சான்றிதழ் ,CC - சாதிச் சான்றிதழ் ,PC - பட்டா நகல் ,GP - புகார்; மனு

இந்தக் குறுஞ்செய்தி, தேசிய தகவலியல் மையத்தின் வழங்கியில் ஒரு மின்னஞ்சலாகச் சேமிக்கப்படும் . செய்தி அனுப்பியவருக்குப் பெற்றுக்கொண்டதற்கான உறுதிச் செய்தியும் அனுப்பப்படும் .இந்தக்

குறுஞ்செய்தி, தொடர்புடைய அலுவலரின் அலுவல்சார் மின்னஞ்சலுக்கு உடனே மடைமாற்றப்படும் .அந்த அலுவலர், அந்த மின்னஞ்சலை அன்றன்றே தரவிறக்கி, அச்சிட்டு, கோப்புகளில் சேர்ப்பார் .அந்த அச்சப் பிரதி, வழக்கமாகக் காகிதத்தில் அளிக்கப்படும் மனுவைப் போன்றே கருதப்படும் .

திருவள்ளூர் மாவட்டத்தின் சேவை அளிப்பு முறையின்படி, குறுஞ்செய்தியின் மின்னஞ்சல் வடிவம், அச்சப் படியாகச் சேமிக்கப்படுகிறது .இதனால், மனுதாரர் தரப்பில் நேரம், பணம், ஆற்றல் ஆகியவை சேமிக்கப்படுகின்றன .ஆனால், அரசு அலுவலகத் தரப்பில் காகிதச் சமை தொடர்கிறது .இது, காகிதமில்லா அலுவலகம் என்ற இலக்கினைக் கொண்டுள்ள மின்னாளுகை நடைமுறைக்கு எதிரானது .எனவே, அச்சப் படி எடுக்காமல் இந்தச் சேவையை வழங்கும் வகையில் இந்தத் திட்டத்தை மறு ஆய்வு செய்திடல் வேண்டும்.

3. புதிய செல்மொழி

செல்பேசிகளில் குறுஞ்செய்திகள் மூலமாகவும் மின்னரட்டை வாயிலாகவும் புதிய குறுமொழி உருவாகியுள்ளது .இது, முழுதும் ஆங்கிலத்திலேயே உள்ளது .ஒருவரின் சொற்குறுக்கத்தைப் பலரும் தொடர்ந்து பயன்படுத்துகின்றனர் .இதனால், குறிப்பிட்ட சொற்குறுக்கம், குறுமொழியினுள் புதிய சொல்லாக நிலைத்துவிடுகிறது .இவை மின்னரட்டையிலும் பயன்படுத்தப்படுகின்றன .இவை குறைவான நேரத்தில் வேகமாகக் கலந்துரையாட உதவுகின்றன .

தனிக் களஞ்சியமாகவே மாறிவிட்ட சொற்குறுக்கங்களின் பட்டியலில் அதிகமாகப் பயன்படுத்தப்படும் சில இங்கே தரப்பட்டுள்ளன:¹⁷

ஆங்கிலச் சொற்குறுக்கங்கள்			
சொற்குறுக்கம்	முழு வடிவம்	சொற்குறுக்கம்	முழு வடிவம்
ASL	Age, Sex, Location	RUOK	Are you OK?
ASAP	As soon as possible	TIA	Thanks in advance
BBL	Be back later	J/K	Just kidding
BRB	Be right back	TTFN	TaTa for now
LOL	Laughing out loud	BFN	Bye for now
ROFL	Rolling on the floor laughing	Y	Why
BTW	By the way	5n	Fine
OIC	Oh, I see	Gr8	Great
CUL	See you later	Gd n8	Good Night
OTOH	On the other hand	Gd Mg	Good Morning
GMTA	Great minds think alike	1 4 3	I Love You
IMHO	In my humble opinion	FAQ	Frequently Asked Questions

தனியார் மட்டுமின்றி, அரசும் இத்தகைய சொற்குறுக்கங்களை அலுவல்பூர்வமாக உருவாக்கி வருகிறது . திருவள்ளூர் என்பதை TLR என்றும், கன்னியாகுமரி என்பதை KKM என்றும் அரசு குறிப்பிடுகிறது . குறுஞ்செய்திச் சேவை முழுவதும் இத்தகைய சொற்குறுக்கங்களைக் கொண்டே நிகழ்கின்றன .இவற்றை உலக அளவிலும் இந்திய அளவிலும் தரப்படுத்தல், மிக முக்கியம்.

இந்தப் புதிய குறுமொழியில் தமிழின் பங்கினை உறுதிப்படுத்த வேண்டிய தேவை உள்ளது .தமிழில் புதிய சொற்குறுக்கங்களை உருவாக்கி, பரவலாகவும் ஆழமாகவும் பயன்படுத்திடல் வேண்டும் .ஆங்கிலத்தில் உள்ள சொற்குறுக்கங்களுக்குத் தமிழில் மாற்றுக் குறுக்கங்களை உருவாக்க வேண்டும் .மேலும் , ஆங்கிலத்தில் வருவதற்கு முன்பாகவே, தமிழில் புதிய சொற்குறுக்கங்களை உருவாக்கி, தனிப் பாதை வகுத்திடல் வேண்டும் .

இதற்கு முதற்படியாகத் தமிழில் குறுஞ்செய்திகளை அனுப்பும் வசதியை அனைத்துச் செல்பேசிகளிலும் அறிமுகப்படுத்த வேண்டும் .விசைப்பலகைத் தரப்படுத்தலும் அதை நடைமுறைப்படுத்தலும் இதற்கு மிக முக்கியம் ஆகும்.

தமிழில் உள்ள சொற்குறுக்கங்கள்

தமிழக அரசு மருத்துவமனைகள் வழங்கும் மாத்திரைகளில் தமிழக அரசு என்பதைத் 'த/அ' எனக் குறிப்பிடுகின்றனர் .அரசாணைகளில் 'த.நா.' என்ற குறுக்கம் பயன்படுகிறது .குடும்பக் கட்டுப்பாடு என்பதைக் குறிக்கும் 'கு.க.' என்ற குறுக்கமும் புகழ்பெற்றது .

தமிழகத்தில் உள்ள ஆவடி (AVADI) என்ற நகரமே "Armoured Vehicles and Ammunition Depot of India" என்ற ஆங்கில வாக்கியத்தின் சுருக்கம் ஆகும் .கலைஞர் கருணாநிதி நகர் என்பது கே.கே .நகர் என்றும் ஜெ.ஜெயலலிதா நகர் என்பது, ஜெ.ஜெ .நகர் என்றும் சுருங்கியுள்ளன .ஆங்கில வழியில் பற்பல சுருக்கங்கள், தமிழில் வேகமாக நிலைத்து வருகின்றன .இந்நிலையில் தமிழ் வாயிலான சொற்குறுக்கங்கள், அரிதாகவே காணப்படுகின்றன .தமிழக அரசியல் துறையில் பயன்படுத்தப்பெறும் சொற்குறுக்கங்களுள் சில:

தமிழில் அரசியல் துறையில் உள்ள சொற்குறுக்கங்கள் சில	
அ.இ.அ.தி.மு.க.	அகில இந்திய அண்ணா திராவிட முன்னேற்றக் கழகம்
காங்.	காங்கிரஸ்
கொ.ப.செ.	கொள்கை பரப்புச் செயலாளர்
ஜெ.	ஜெயலலிதா
த.மு.மு.க.	தமிழ்நாடு முஸ்லிம் முன்னேற்றக் கழகம்
தி.க.	திராவிடர் கழகம்
தி.மு.க.	திராவிட முன்னேற்றக் கழகம்
திருமா	திருமாவளவன்
து.செ.	துணைச் செயலாளர்
தே.மு.தி.க.	தேசிய முற்போக்கு திராவிடக் கழகம்
பா.ம.க.	பாட்டாளி மக்கள் கட்சி
ம.தி.மு.க.	மறுமலர்ச்சி திராவிட முன்னேற்றக் கழகம்
ம.பொ.சி.	ம.பொ.சிவஞானம்
மா.செ.	மாவட்டச் செயலாளர்
மு.க.	மு.கருணாநிதி
மூ.மு.க.	மூவேந்தர் முன்னேற்றக் கழகம்
வ.செ.	வட்டச் செயலாளர்
வி.சி.க.	விடுதலை சிறுத்தைகள் கட்சி
வைகோ	வை.கோபால்சாமி

தமிழில் தோன்றியுள்ள சொற்குறுக்கங்களை முழுமையாகப் பட்டியலிடல் வேண்டும்; இவற்றைத் தரப்படுத்தல் வேண்டும்; மேலும் புதிய சொற்குறுக்கங்களைப் பெருமளவில் உருவாக்க வேண்டும் .பின்னர் செல்பேசி, இணையம், நாளிதழ், வார இதழ், தொலைக்காட்சி, வானொலி உள்பட அனைத்து ஊடகங்கள் வாயிலாகவும் இவற்றைப் பரப்புதல் வேண்டும் .அப்பொழுதுதான் நிகழ்காலத் தேவைகளை நிறைவு செய்ய இயலும்; மொழி ரீதியில் எதிர்காலப் போட்டிகளையும் சமாளித்து, நிலைத்து நிற்க இயலும்.

4. செல்பேசியில் தமிழ் -சில முன்முயற்சிகள்

செல்பேசியில் தமிழ்ப் பயன்பாட்டினை அதிகரிக்க, உலகெங்கும் பல முன்முயற்சிகள் எடுக்கப்பட்டுள்ளன . அவற்றுள் சில இங்கே.

செல்லினம் -தமிழ்க் குறுஞ்செய்திச் சேவை

செல்லினம் ,ஒருங்குறியைப் பயன்படுத்தி, செல்பேசிகளில் தமிழ்க் குறுஞ்செய்திகளை அனுப்பும் மென்பொருளாகும் .இதனை மலேசியாவைச் சேர்ந்த முத்து நெடுமாறன் தலைமையிலான முரசு நிறுவனம், 2003இல் உருவாக்கியது .பின்னர், தைப்பொங்கல் திருநாளான ஜனவரி 15, 2005 முதல் சிங்கப்பூரில் வணிகப் பயன்பாட்டுக்கு வெளியானது. 2005ஆம் ஆண்டு “Most Innovative Mobile Application” என்ற பிரிவில், மலேசிய அரசின் ஆதரவில் வழங்கப்படும் “Malaysian ICT Excellence Award” என்ற விருதினைச் செல்லினம் வென்றது .பல வகையான செல்பேசிகளில் பயணித்த செல்லினம், 2009ஆம் ஆண்டு ஜூன் மாதத்திலிருந்து முதல் ஐ-ஃபோனிலும் இயங்கி வருகிறது.¹⁸

செல்பேசியில் நூல்கள் -ஒலி நூல்கள்

மொபைல்வேதா (www.fublish.com) நிறுவனத்தைச் சேர்ந்த கணேஷ்ராம் என்பவர் ,செல்பேசியில் நூல்களையும் ஒலி நூல்களையும் உருவாக்கியுள்ளார் .'சீட்' (SEED) என்ற பெயரில் ஒரே நுண்சில்லில் ஆயிரம் நூல்களை உள்ளடக்கினார் .இதில் தொல்காப்பியம், திருக்குறள், பத்துப் பாட்டு, எட்டுத் தொகை, கம்பராமாயணம் முதல் தற்கால இலக்கியம் வரையிலான 600 தமிழ் நூல்களும் 400 ஆங்கில நூல்களும் இருந்தன .பத்து ஒலி நூல்களும் இருந்தன .இந்த நுண்சில்லை 27. 01.2009 அன்று வேலூரில் வி.ஐ.டி . பஸ்கலைக்கழகத்தில் அதன் வேந்தர் ஜி.விஸ்வநாதன் வெளியிட்டார் .இத்தகைய முயற்சி, இந்தியாவிலேயே தமிழில்தான் முதல் முறையாக நடந்துள்ளது.¹⁹

செல்பேசி நூல்களை வேண்டிய இடங்களில் வேண்டிய நேரத்தில் படிக்க முடியும் .பார்வைத் திறனுக்கு ஏற்ப, எழுத்துகளின் அளவைப் பெரிதாக்கிப் படிக்கலாம் .எழுத்துகளின் நிறங்களையும் மாற்றலாம் . விரலால் மீண்டும் மீண்டும் அழுத்தி, அடுத்தடுத்த பக்கங்களுக்குப் போக வேண்டுமே என்ற சோர்வும் வேண்டாம் .நாம் விரும்பும் நேர அளவில், தானே பக்கங்கள் மாறும் வசதியும் உண்டு .படித்த இடத்தை நினைவிற்கொள்ள ,பக்கக் குறிப்பீடு)புக் மார்க் (வசதியும் உண்டு .பாட்டுக் கேட்பது போல் நூலைக் கேட்கவும் முடியும் .

தமிழில் செல்பேசி விசைப்பலகைத் தரப்படுத்தல்

சில செல்பேசி நிறுவனங்கள், தமிழில் குறுஞ்செய்தி அனுப்ப, அவரவர்க்கென ஒரு விசைப்பலகையை வடிவமைத்துப் பயன்படுத்தி வருகின்றனர்.²⁰ இதனால் செல்பேசி நிறுவனங்களுக்கும் பயனர்களுக்கும் பல சிக்கல்கள் எழுகின்றன .இதனால் செல்பேசிகளில் தமிழ் விசைப் பலகையைத் தரப்படுத்த வேண்டிய தேவை எழுந்தது .இதற்காகத் தமிழக அரசு, 28.03.2007அன்று, தகவல் தொழில்நுட்பத் துறையின் சார்பில், G.O. Ms.No. 10எண்ணிட்ட அரசாணை மூலம் முனைவர் பொன்னவைக்கோ தலைமையில் ஒரு பணிக் குழுவை அமைத்தது .அதில் சி.உமாசங்கர்)முன்னாள் மேலாண் இயக்குநர், எஸ்காட் ,(முனைவர் கிருஷ்ணமூர்த்தி)கணித் தமிழ்ச் சங்கம் ,(கிருஷ்ணன்)பி.எஸ்.என்.எல் ,(ஏ.இளங்கோவன்)உத்தமம் ,(முனைவர் ப.அர.நக்கீரன்)இயக்குநர், தமிழ் இணையப் பல்கலைக்கழகம் (ஆகியோர் இடம் பெற்றனர்.

இந்தச் செல்பேசி விசைப்பலகை தரப்பாட்டுக் குழு, ஏழு முறைகள் கூடி, பல்வேறு கருத்துகளையும் சோதனை முடிவுகளையும் ஆராய்ந்தது .இறுதியாக 01.04.2008அன்று கூடி, ஆய்ந்தது .அதன் பயனாக , பல்-முறை தட்டல் ,இரு-விசை தட்டல் ,மெய்-உயிர் தட்டல் ஆகிய பயன்பாட்டு முறைகளைப் பரிந்துரைத்தது.²¹ இந்தப் பயன்பாட்டு முறைகளில் ஏதேனும் ஒன்றினை, தரப்படுத்தப்பட்ட விசைப்பலகை வடிவமைப்புடன் செல்பேசி நிறுவனங்கள் செயல்படுத்திக்கொள்ளலாம் என இக்குழு அறிவித்துள்ளது.

தரப்பாட்டுக் குழு பரிந்துரைத்த தமிழ் விசைப்பலகை:

1 க ங அ ஆ	2 ச ஞ இ ஈ	3 ட ண உ ஊ
4 த ந எ ஏ	5 ப ம ஐ	6 ய ர ஒ ஓ
7 ல வ ஒள	8 ழ ள ஃ	9 ற ன @
* DELINK	0 SPACE	# . ஸ...பூநீ

இந்த விசைப்பலகையைப் பயன்படுத்தும் முறைகளையும் அக்குழு விவரித்துள்ளது .இந்தப் பரிந்துரையைப் பரிசீலித்து, அரசு உடனடி நடவடிக்கை எடுக்க வேண்டும்.

செல்பேசி துறைக்கான கலைச்சொற்கள்

'செல்பேசி தொழில்நுட்பத் தமிழாக்கமும் தரப்படுத்தலும்' பற்றிய தேசிய மாநாடு, 2006 ஜூலை 22 அன்று சென்னையில் நடைபெற்றது .அதற்கான மாநாட்டுக் கட்டுரைகள் தொகுப்பில், 'செல்பேசி துறைக்கான கலைச்சொற்கள்' என்ற பகுதி இடம் பெற்றுள்ளது .இதில் செல்பேசித் துறையிலும் கணினி, இணையத் துறைகளிலும் பயன்படுத்தப்படும் ஆங்கிலச் சொற்கள் பலவற்றுக்குத் தமிழில் இணையான சொற்கள் தொகுக்கப்பெற்றுள்ளன .இவற்றைக் கணித் தமிழ்ச் சங்கத் தலைவர் மா.ஆண்டோ பீட்டர் தொகுத்தார் . அந்தத் தொகுப்பிலிருந்து சில சொற்கள் இங்கே:

செல்பேசி துறைக்கான கலைச்சொற்கள்	
Arrow key	திசை விசை
Audio Response Device	கேட்பொலி மறுமொழிச் சாதனம்
Auto Dial	தானியங்கு சுழற்றி
Auto Text	உடனடி உரை
Back Space	பின்னிடம் (ஒரு விசை)
Bandwidth	கற்றை அகலம்

BIOS - Basic Input / Output System	அடிப்படை உள்ளீட்டு / வெளியீட்டு முறைமை
Cache Memory	இடைமாற்று நினைவகம்
Caller ID	அழைத்தவர் அடையாளம்
Card Reader	அட்டை படிப்பி
Expanded Memory	விரிவாக்க நினைவகம்
Icon	சின்னம்
Missed call	பேசத் தவறிய அழைப்புகள்
Remote access	தொலைநிலை அணுக்கம்

செல்பேசியில் தமிழ்த் தளங்களை வாசிக்க

செல்பேசிகளில், தமிழ்த் தளங்கள் சரிவர தெரிவதில்லை என்ற குறைபாடு இருந்தது .இந்தக் குறையை ஒபெரா மினி உலாவி (<http://www.opera.com/mini>) தீர்க்கிறது.²² ஆயினும் இது, ஒருங்குறியில் இயங்கும் தளங்களை நன்கு காட்டுகிறது .ஆனால், இதர குறிமுறைகளிலும் எழுத்துருக்களிலும் உள்ள தளங்களிலிருந்து எழுத்துருக்களைச் செல்பேசியில் நிறுவ இயலவில்லை .இச்சிக்கலைத் தீர்க்க, செல்பேசி உற்பத்தி நிறுவனங்கள் முயல வேண்டும்.

5. செல்லிட ஆளுகையில் சிக்கல்களும் வாய்ப்புகளும்

செல்பேசிகள் பன்முகத் தன்மை கொண்டவை .அவற்றின் வழியே பல்வகைச் செயல்களை நிகழ்த்த முடியும் . ஆயினும், குறைந்தபட்சம் ,செல்பேசிகள் வழியே மூன்று வழிகளில் செல்லிட ஆளுகை நடைபெறுகிறது . அவை:

- செல்பேசி வழியே இணையத் தொடர்பு
- அரசுடன் குறுஞ்செய்திப் பரிமாற்றம்
- அரசு அலுவலருடன் நேரடி உரையாடல்

செல்பேசி வழியே இணையத் தொடர்பு:

தமிழக நிலையை வைத்துப் பார்க்கையில், செல்பேசி வழி இணையத் தொடர்பு, மிகக் குறைவாகவே பயன்பாட்டில் உள்ளது .இத்தகைய வசதி கொண்ட சிலரும், அதனை அரசுத் தளங்களைப் பார்ப்பதற்கு அதிகம் பயன்படுத்துவதில்லை .இன்னொரு கோணத்தில், செல்பேசிவழியே பார்ப்பதற்கு ஏற்ப, அரசு இணையத்தளங்கள் எவையும் தனித் தளத்தினை உருவாக்கவில்லை .ஆயினும் எதிர்காலத்தில் இந்த முறை, அதிகப் பயன்பாட்டுக்கு வரும் என எதிர்பார்க்கலாம்.

அரசுடன் குறுஞ்செய்திப் பரிமாற்றம்:

தற்பொழுதைய நிலையில் அரசுடன் குறுஞ்செய்தி பரிமாற்றம், ஆங்கில வழியாகவே நிகழ்கிறது .தமிழில் குறுஞ்செய்தி அனுப்புவதில் உள்ள தடைகளைக் களைந்திடல் வேண்டும் .விசைப்பலகைத் தரப்படுத்தல், கலைச்சொற்கள், சொற்குறுக்கங்கள் ஆகிய தேவைகளை முழுமையாகவும் வேகமாகவும் நிறைவேற்றினால் மட்டுமே தமிழ்க் குறுஞ்செய்திப் பயன்பாடு பரவும் .அதன் பின்னரே அரசுச் சேவைகளை அதன்வழி வழங்க முடியும்.

அரசு அலுவலருடன் நேரடி உரையாடல்:

தமிழக அரசின் பல்வேறு துறை அலுவலர்களின் செல்பேசி எண்கள், வெளிப்படையாக அறிவிக்கப்பட்டுள்ளன .நாளேடுகளிலும் இணையத்தளங்களிலும் இவை உள்ளன .ஆனால், அனைத்து

அலுவலர்களின் எண்களும் அறிவிக்கப்படவில்லை .அது மட்டுமின்றி, சில தருணங்களில் அரசு அலுவலர்களின் செல்பேசி எண்களைத் தொடர்பு கொண்டால், அவை அணைக்கப்பட்டுள்ளன; தொடர்பு எல்லையில் இல்லை என்பன போன்ற அறிவிப்புகள் ஒலிக்கின்றன .இத்தகைய சமயங்களில் அழைத்தவரின் குரலைப் பதிந்து சேமிக்க வேண்டும். பின்னர், அலுவலர், அழைத்தவரைத் தொடர்புகொண்டு குறைகளைக் கேட்டுத் தீர்வளிக்க வேண்டும் .இதற்கு வாய்ஸ்நாப் (<http://www.voicesnap.com>) போன்ற மென்பொருள்கள் பயன்படும்.

செல்பேசியில் அரசு அலுவலருடன் பேசித் தகவல்களைப் பெறுவதை அலுவலர்கள் சிலர் விரும்பவில்லை . இது மரியாதைக் குறைவான செயல் என்றும் மக்கள் நேரில் வந்து கேட்க வேண்டும் என்றும் அவர்கள் நினைக்கிறார்கள் .இந்த மனோபாவம், மின்னாளுகைக்கும் செல்லிட ஆளுகைக்கும் தடையாக உள்ளது . இதைக் களைய, அரசு அலுவலர்களுக்குத் தனிப் பயிற்சி அளிக்க வேண்டும்.

இந்தியாவில் செல்பேசிவழி இணையப் பயன்பாடு

இந்தியாவில் செல்பேசிகள் வழியே இணையத்தைத் தொடர்ச்சியாகப் பயன்படுத்துவோரின் எண்ணிக்கை, சுமார் 20 இலட்சம் மட்டுமே .இங்கே தொடர்ச்சியான பயனாளர் என்ற பதம், மாதத்திற்கு ஒரு முறை இணையத்தைப் பயன்படுத்தியவரையே குறிக்கிறது .இந்திய இணையம் & செல்பேசிக் கழகமும்)IAMAI - Internet & Mobile Association of India) ஐ.எம்.ஆர்.பி.யும் இணைந்து நடத்திய ஆய்வில் இது தெரிய வந்துள்ளது.²³

இந்தியாவில் உள்ள செல்பேசிகளுள் இணைய வசதி கொண்டவை, 12.7 கோடி .இவர்களுள் ஆண்டுக்கு ஒரு முறையேனும் செல்பேசி வழியே இணையத்தை அணுகியவர்களின் எண்ணிக்கை, ஒரு கோடியே 20 இலட்சம் .இவர்களுள் மாதத்திற்கு ஒரு முறை இணையத்தை அணுகியோர், 20இலட்சம் மட்டுமே.

இந்த 20 இலட்சம் பேர்களில் 18-35 வயதுக்குள் உள்ளோர், 70% பேர் .இவர்கள், பெரும்பாலும் கல்லூரி மாணவர்களும் இளைஞர்களும் ஆவர் .இணையத்தைப் பயன்படுத்துவோரில் 60% பேர்கள், மின் அரட்டைக்கும், நுண் வலைப்பதிவு, சமூகப் பிணையத் தளங்கள் ஆகியவற்றுக்கும் பயன்படுத்துகின்றனர் . அடுத்து, 23% பேர்கள், தகவல் தேடலுக்காக இணையத்தை நாடியுள்ளனர் .இந்தியா முழுவதும் 40 ஆயிரம் பேர் மட்டுமே மின் வணிகத்திற்காக, செல்பேசி வழியே இணையத்தை அணுகியுள்ளனர்.

இந்தப் புள்ளி விவரத்தின் வாயிலாக, செல்பேசி வழியே இணையத்தை அணுகுவதில் பல சிக்கல்களும் தடைகளும் உள்ளன என அறியலாம் .இணைய உலாவல் கட்டணம், இணைய உலாவல் வேகம், சிறிய திரையில் இணையத்தளங்களைப் பார்ப்பதில் உள்ள சிரமம், நச்சு நிரல்கள் தாக்கும் அச்சம் ...உள்ளிட்ட இதற்குக் காரணங்களாய் இருக்கலாம் .இது, தனித்த ஆய்வுக்கான களமாகும்.

செல்பேசி எண்களைத் திரட்ட வேண்டும்

அரசு முதல் கட்டமாக, செல்பேசி எண்ணைத் தெரிவிக்க ஆர்வமுள்ள மக்கள் அனைவரின் எண்களையும் திரட்ட வேண்டும் .குடும்ப அட்டை வைத்திருப்போர், வரி செலுத்தவோர், அரசு ஊழியர், ஆசிரியர்கள் என வகை வாரியாக செல்பேசி எண்களைத் திரட்டிப் பட்டியலிட வேண்டும் .அதன் பின்னர், அந்தந்த வகையினருக்கு அவர்களுக்கு ஏற்ற செய்திகளையும் சேவைகளையும் குறுஞ்செய்திகளாக அனுப்பலாம் . அரசின் பல்வகைத் திட்டங்களையும் சலுகைகளையும் கட்டண நினைவூட்டல்களையும் அறிவிப்புகளையும் குறுஞ்செய்திகளாக அனுப்பலாம்.

குறுஞ்செய்தி வழியே முன்னறிவிப்புகள்

அடுத்த நாள் மின் விநியோகம் தடைப்படும் பகுதிகளில் உள்ள மக்கள் அனைவருக்கும் 24மணி நேரத்திற்கு முன்பே, குறுஞ்செய்தி அனுப்பிவிடலாம் .இதன் மூலம் மக்கள், தகுந்த முன்னேற்பாடுகளைச் செய்ய ஏதுவாகும் .இப்போது, நாளிதழ்கள், வானொலி ஆகியவற்றில் இந்தச் செய்திகள் வெளியாகின்றன .

ஆயினும் செய்திகள் அனைவரையும் சென்றடைவதில்லை .அதே போன்று, நியாய விலைக் கடைகளில் எந்தெந்த நாட்களில் என்னென்ன பொருட்கள் விநியோகிக்கப்படுகின்றன என்பதையும் முன்கூட்டியே அறிவிக்கலாம் .குறிப்பிட்ட நாளில் வழங்க வேண்டிய பொருள் இருப்பு இல்லையெனில், அதையும் நுகர்வோருக்கு முன்னதாகவே தெரிவிக்கலாம் .இதன் மூலம் 'இன்று மண்ணெண்ணெய் இல்லை', 'சர்க்கரை இல்லை' என மக்கள் வெறுங்கையுடன் திரும்பி வருவதைத் தவிர்க்கலாம். இவ்வாறே அனைத்துத் துறைகளும் சேவைகளை அளிக்க முடியும்.

ஒரே நேரத்தில் பேரெண்ணிக்கையில் குறுஞ்செய்திகள் அனுப்பிட, பல்வகை மென்பொருட்களும் திட்டங்களும் உள்ளன .இதற்கு மேப்டெக் இன்போடெக் (<http://www.maptechindia.com>) நிறுவனத்தின் மென்பொருட்களையும் பரிசீலிக்கலாம். இணையம் வழியாகவும் பேரெண்ணிக்கையில் குறுஞ்செய்திகள் அனுப்பிட இயலும் .இத்தகைய சேவைகளைப் பல நிறுவனங்கள், இலவசமாகவும் வழங்கி வருகின்றன . இவற்றையும் அரசுத் துறைகள் பயன்படுத்தலாம்.

செல்பேசிகள் அளிக்கும் எல்லையற்ற வாய்ப்புகள்

செல்பேசியில் ஒருவரை ஒருவர் பார்த்தபடியே பேசுவதற்கும் வசதி உண்டு .இதைப் பயன்படுத்தி மக்கள், அலுவலரையும் அமைச்சர்களையும் நேருக்கு நேர் பார்த்துப் பேசிட வழி காண வேண்டும் .

குறுஞ்செய்தி போலவே படச் செய்திகளையும் காணொலிப் படக் காட்சிகளையும் அரசுக்கு அனுப்பிட, மக்களைப் பயிற்றுவிக்க வேண்டும் .ஓரிடத்தில் நிகழும் விபத்து, சட்ட மீறல், கையூட்டு உள்ளிட்ட காட்சிகளை காணொலிப் படமாக எடுத்து அரசுக்கு அனுப்பக் கோரலாம் .இவ்வகைக் காட்சிகளைப் பெற்று, மேல் நடவடிக்கை எடுத்திட, தொழில்நுட்ப வகையிலும் மனப்பாங்கு அடிப்படையிலும் அரசு ஆயத்தமாய் இருக்க வேண்டும்.

குறுஞ்செய்திச் சேவைகளுக்காகத் தமிழக அரசின் ஒவ்வொரு துறையும் தனித் தனி எண்களை அறிவிக்கத் தேவையில்லை .இதற்குப் பதிலாக, ஒரே எண்ணை அறிவித்து, அதன் வழியே ஒருங்கிணைந்த சேவைகளை அளிக்க வேண்டும் .அரசுக்குச் செலுத்த வேண்டிய கட்டணங்களைச் செல்பேசிகள் வழியே செலுத்துவற்கு ஏற்ப, மக்களைப் பயிற்றுவிக்க வேண்டும் .

உரையாடல், குறுஞ்செய்தி, இணையத் தொடர்பு ஆகியன மட்டும் இல்லாமல், செல்பேசிகளின் முழுத் திறனையும் பயன்படுத்தும் வண்ணம் செல்லிட ஆளுகையைக் கட்டமைக்க வேண்டும் .மக்களாட்சியின் முழுமுதற்பயனே மக்கள் அதிகாரம் பெறுவதில்தான் உள்ளது .எல்லையற்ற வாய்ப்புகளை அளிக்கும் செல்பேசிகள் வழியே கடைக்கோடி மனிதனும் அதிகாரம் பெற வாய்ப்பு உண்டு .பாமர மனிதனின் தேவைகளை அவன் கேட்கும் முன் அளிக்கும் உயர்ந்த நிர்வாகத்தை எட்ட, செல்லிட ஆளுகை உறுதியாக உதவும்.

அடிக்குறிப்புகள்

1. Telecom Subscription Data as on 31st March 2010, Telecom Regulatory Authority of India (TRAI), <http://www.trai.gov.in/WriteReadData/trai/upload/PressReleases/732/pr26apr10no20.pdf>
2. Nokia to introduce mobile phones Rs.500 in india, http://www.knowyourmobile.in/news/413555/nokia_to_introduce_mobile_phones_rs_500_in_india.html
3. செல்போன் கோபுரங்களுக்கு வரி :சென்னை மேயர், நக்கீரன், <http://www.nakkheeran.in/users/frmNews.aspx?N=17666>
4. Mobile Number Portability -Launch Date -May 1st week in Chennai Bangalore – I&B Minister Raja, <http://www.moneymint.in/mobile/mobile-number-portability-launch-date-may-1st-week-in-chennai-bangalore-ib-minister-raja>

5. Johan Hellstrom, Mobile phones for good governance - challenges and way forward,
http://www.w3.org/2008/10/MW4D_WS/papers/hellstrom_gov.pdf
6. IRCTC announces ngpay as the leading mobile sales channel for purchasing Rail tickets,
<http://www.ngpay.com/site/pressroom2.htm>
7. tnresults.nic.in SSLC Results 2009 10th Results of Tamilnadu,
<http://ready2beat.com/educational/tnresultsnicin-sslc-results-2009-10th-results-tamilnadu>
8. Mobile-based property tax payment system launched in Chennai, Chennai Online,
<http://news.chennaionline.com/newsitem.aspx?NEWSID=20f3a55e-3520-413d-a72c-77affb388d56&CATEGORYNAME=CHN>
9. புதிய ரேஷன் கார்டு தயாரா? எஸ்.எம்.எஸ் .அல்லது இமெயில் மூலம் தெரிந்துகொள்ளலாம், நக்கீரன்,
<http://www.nakkheeran.in/users/frmNews.aspx?N=24681>
10. இனி எஸ்எம்எஸ் மூலம் பென்ஷன் பற்றிய தகவல்!, தடஸ்தமிழ்,
<http://thatstamil.oneindia.in/news/2009/12/03/sms-service-pensioners.html>
11. SMS your traffic complaints, The Hindu,
<http://www.hindu.com/2006/07/21/stories/2006072118830300.htm>
12. போக்குவரத்து நெரிசல் குறித்து எஸ்.எம்.எஸ் .மூலம் அறிவிக்க புதிய திட்டம் :கூடுதல் கமிஷனர் ஷகீல் அக்தர் தகவல், http://news.tninus.org/index.php?option=com_content&view=article&id=3934:2009-11-03-06-43-14&catid=268:poverty-alleviation-&Itemid=334
13. சுற்றுலா விடுதிகளை எஸ்.எம்.எஸ்.ல் முன்பதிவு செய்யலாம்,
<http://www.erodelive.com/entertainment/news.php?id=432>
14. காவல்நிலையத்திற்கு எஸ்.எம்.எஸ் மூலம் புகார்,
<http://www.chennaionline.com/tamil/news/newsitem.aspx?NEWSID=9a91b3a5-6221-45ca-a143-7d7876b3167c&CATEGORYNAME=TNATL>
15. Deepa H Ramakrishnan, Parents to get SMS about children skipping classes in schools,
<http://www.hindu.com/2009/07/30/stories/2009073059040300.htm>
16. நாட்டிலேயே முதல் முறையாக அறிமுகம் எஸ்.எம்.எஸ்., மூலம் கலெக்டருக்கு கோரிக்கை, தினமலர்,
http://www.dinamalar.com/General_detail.asp?news_id=8542
17. அரட்டை)ம (எஸ்.எம்.எஸ் .விளங்கா சொற்கள், <http://www.cert-in.org.in/securepc/ta/resources/thingstoknow/acronyms.html>
18. செல்லினம், http://sellinam.com/?page_id=2
19. அண்ணாகண்ணன், செல்பேசிக்குள் ஒரு நூலகம்,
<http://chennaionline.com/tamil/tamilcolumn/newsitem.aspx?NEWSID=28dd83b7-2753-40aa-9242-afbe9316e571&CATEGORYNAME=Anna>
20. Rohan Samarajiva, Tamilnadu adopts Tamil SMS solution developed in Sri Lanka,
<http://lirneasia.net/2006/05/tamilnadu-adopts-tamil-sms-solution-developed-in-sri-lanka/>
21. தமிழில் செல்பேசி விசைப்பலகை வடிவமைப்பு -தரப்படுத்தலும் செயலாக்கமும்,
http://www.tamilvu.org/coresite/download/STMP_Report_Tamil.pdf
22. மொபைல் ஃபோனில் தமிழ் தளங்களை வாசிக்க, <http://inneram.com/200909033856/how-to-read-tamil-unicode-sites-from-wifi-phones>.
23. 2 million serious Mobile Internet users,
http://www.iamai.in/PRelease_detail.aspx?nid=2000&NMonth=1&NYear=2010

Thirukkural Mobile

A Cultural Tool for All

G. Bhuvan Babu, MFA.

bhuvan@avonmobility.com

ph - 98416 81233

Introduction

Thirukkural is not an imaginary utopia or a heaven for human beings to aspire for. It is out and out concerned with today's world. Over the centuries Thirukkural has become a much-quoted, quintessential guidebook for life. They are heard in the speeches of ministers, seen written above the driver seat on most of the buses that range Tamil Nadu and small children memorize the Kural in order to be able to chant it verse after verse - many can recite the entire 1,330 verses by heart.



But still, is this masterpiece only to teach vaguely in schools for the sake of memorising and to be used as quotations by speakers? No. Not at all. It was written to enrich human values. Thiruvalluvar's intuition and insight is solely concerned with human life in the present world.

Gandhiji was one of those, who realized the greatness of Thirukkural and its need for enlightenment in the society. He said: "I learnt Tamil only to enable me to study Thiruvalluvar's Kural through his mother tongue itself". He further said: "Only a few of us know the name of Tiruvalluvar. The North Indians do not know the name of the great saint. There is none who has given such a treasure of wisdom like him."

This statement was given six decades before and it still remains true. Though many of us know Thirukkural and thiruvalluvar, only few people realize its values and try to follow it in day to day life in this Tamil speaking land.

Objective

The objective of this paper is to reach/teach this cultural treasure to the common man in the best possible way

- with a greater impact on its relevance and significance in this modern world
- in a simple, understandable language, making learning interesting and fun
- to apply and practice its values and ethics in day to day business and life style situations.

Research & Results

Before finding ways to achieve the objectives, it is very essential to acknowledge our Government and Tamil Scholars for their great efforts to create awareness of Thirukkural in the common man. We worship Thiruvalluvar as a great saint, the universal teacher of our lives. Valluvar Kottam is erected with all 1330 couplets inscribed in it. Each year, on Thiruvalluvar day, we celebrate him with a public holiday. We have erected a beautiful shrine to him in the midst of a garden in Mylapore, every year celebrating a great

festival there. The greatest physical monument, however, is the magnificent 133-foot high statue of the saint stood majestically off the shores of Kanayakamari, the most southern point in India standing like some spiritual lighthouse watching over the world.

With the painstaking efforts of Kural Lovers, Thirukkural is also credited to be one of the early adopters of Information Technology. Various attempts have been made to popularize Thirukkural in the form of books, music, e - versions, paintings, radio and television programmes, mobile phones, etc.

Books: Numerous translations and descriptive books are found in the market.

Music: Thirukkural in the form of musical CDs are available.

E-Versions: E-Books, Research papers, blogs and interactive content are found in websites abundantly.

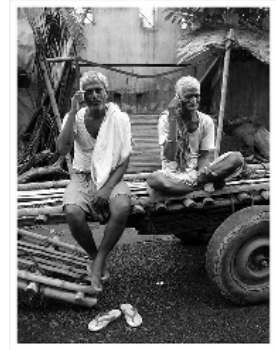
Paintings: Many artists have visualized kurals in the form of paintings. (exhibited at Valluvar Kottam)

Radio and Television: A Kural a day, serials, debates etc.

Mobile Phones: Off late, Thirukkural is introduced into mobile phones in text format.

Since mobile phones are becoming an indispensable gadget irrespective of social status in the present scenario, it would be the right means to achieve our objectives. Let's go through some statistics on mobile phones in India before proceeding further.

Statistics on mobile phones in India shows that rural mobile phone users alone in India cross 100 million this year (2010). With 3G (3rd Generation mobile technology) in India on the front; it is going to be an exciting phase for the rural and urban Indian telecom users. Rural mobile phone users can download content on a click from a link sent, until 3G technology spreads its wings there. The 3G bonanza, which will offer very high-speed mobile wireless services, already has a potential market in India with at least 20 million 3G-enabled phones. Videos can be pushed through video calls which would be a low cost version of watching a video without needing any mobile data plan. This has the capability to reach the masses and is as easy as making a call on your phone. Mobile techno gurus say that the number of people who opt for 3G services would be determined by content.



So, with the entry of 3G mobile technology and multimedia (audio and video) on mobile phones, we believe that Thirukkural will be the best content in the form of education, entertainment and enlightenment at greater heights.

Recommendations

Thiruvalluvar says: "Even though the technical know-how of modern technology is known well, the inherent attitude and the actual prevailing condition of the State should be thoroughly understood and only then the technology should be put to application"(637). With this in mind, we have developed a complete new dimension for Thirukkural on mobile phones. This is a service that helps one understand Thirukkural in depth. It gives a Thirukkural every day through your mobile phone, using animation and audio to make learning interesting and fun. Every Thirukkural Video has a logical sequence of –

- Introduction of Kural,
- Animation and musical rendering of kural to the concept and
- Kural's meaning in Simple Tamil/Hindi followed by the English meaning.



Key Benefits

The animated musical cards provide the intended meaning conveyed by Thiruvalluvar.

Mellifluous music and beautiful animations.

Uneducated rural public can understand the meaning and value of Thirukkural with ease.

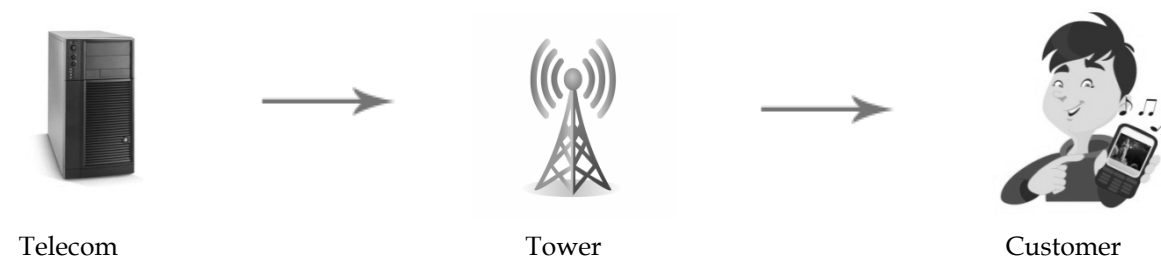
This will certainly be useful not only to Tamil speaking people but also to others to understand the richness of Tamil culture and its heritage.

Also a best friend to the Tamil community, living in various parts of the world, who are not fortunate to educate the new generation in Tamil.

This adds value for Tamil as a classical language (Semmozhi).

How it works?

The customer gets a multimedia Kural every day from the Telecom service providers through a Video Call / link. It is as easy as receiving a call. At one click, the video streams/downloads into the mobile phone and plays a kural with animation, and music, followed by its meaning in a simple language. With the capability to reach the masses through mobile phones and 3G connections, this service is available on a monthly subscription basis for a small fee.



Conclusion

Thiruvalluvar says:

*Resources, means, time, place and deed,
examine these five before you proceed. – 675*

We believe that with Thirukkural itself as the best resource, technology as the means, this is the right time and place to enhance the lifestyle of every individual to create a better world.

செல்பேசிகளில் தரப்படுத்தப்பட்ட தமிழ் இடைமுகம்

மு.சிவலிங்கம்

musivalingam@gmail.com

கோட்டப் பொறியாளர் (ஓய்வு),

பிஎஸ்என்எல் (இந்திய அரசுத் தகவல் தொடர்புத் துறை) – சென்னை.

வேளாண் புரட்சிக்கும், தொழில்புரட்சிக்கும் அடுத்துபடியாகத் தகவல் புரட்சியின் கால கட்டத்தில் நாம் வாழ்கிறோம். வல்லுநர்கள் இப்போதைய கால கட்டத்தைத் 'தகவல் யுகம்' (Information Era) எனக் கணிக்கின்றனர். உலகமெங்கிலும் இன்றைய தகவல் தொடர்பின் வளர்ச்சி இக்கூற்றை உறுதிப்படுத்துவதாக உள்ளது. விரல்நுனியில் தகவல்கள்; விரல்சொடுக்கும் நேரத்தில் தகவல் பரிமாற்றம். பூமிக் கோளத்தில் நாடுகளின் எல்லைகள் மறைந்து போயின. கடல்களின் பரப்பு காணாமல் போனது. தொலைவு என்பது தொலைந்து போயிற்று. உள்ளங்கைக்குள் உலகம் சுருங்கிவிட்டது.

செல்பேசியின் ஊடுருவல்

செல்பேசியின் வருகை தொலைத் தகவல் தொடர்பில் ஒரு புரட்சிக்கு வித்திட்டுள்ளது. இந்தியா உட்பட வளரும் நாடுகளில் செல்பேசியின் வளர்ச்சி புதிய எல்லைகளைத் தொட்டுள்ளது. செல்பேசிப் பயனர்களின் எண்ணிக்கையில் இந்தியா உலகிலேயே மூன்றாவது இடம் வகிக்கிறது. 2010 மார்ச் 1-இல் இருந்த நிலவரப்படி இந்தியாவில் மொத்தத் தொலைபேசிப் பயனர்களின் எண்ணிக்கை 60 கோடிக்கும் சற்றே அதிகம். இது மக்கள் தொகையில் 51.05% ஆகும். இவற்றுள் செல்பேசிப் பயனர்களின் எண்ணிக்கை 56 கோடியே 37 லட்சத்து 30 ஆயிரம். இது மக்கள் தொகையில் 47.91% ஆகும்¹. செல்பேசிப் பயனர்களின் எண்ணிக்கையில் உலகிலேயே சீனாவுக்கு அடுத்தபடியாக இந்தியா இரண்டாவது இடம் வகிப்பது குறிப்பிடத் தக்கது. (சீனாவில் 76 கோடியே 59 லட்சத்து 70 ஆயிரம்).

தமிழ்நாட்டில் 5 கோடியே 22 லட்சத்து 34 ஆயிரம் பயனர்கள் உள்ளனர்². தமிழ்நாட்டின் மக்கள்தொகை அடிப்படையில் இது ஏறத்தாழ 70% ஆகும். அதாவது தமிழ்நாட்டில் நூற்றுக்கு 70 பேர் செல்பேசி பயன்படுத்துகின்றனர். மக்கள்தொகை அடிப்படையில் செல்பேசிப் பயனர்களின் சதவீதத்தில் இந்தியாவிலேயே தமிழ்நாடுதான் முதலிடம் வகிக்கிறது. சின்னஞ்சிறு கிராமத்திலும் செல்பேசிக் கோபுரங்களைக் காண முடிகிறது. படிப்பறிவு இல்லாத மக்கள்கூடச் செல்பேசி பயன்படுத்துகின்றனர். தமிழ்நாட்டில் கேபிள் தொலைக்காட்சி இல்லாத ஊர்களே இல்லை எனலாம். அதற்கு அடுத்தபடியாகச் செல்பேசியைச் சொல்லலாம். தமிழ்நாட்டில் வறுமைக் கோட்டுக்குக் கீழே வாழ்பவர்களும் செல்பேசிப் பயனர்களாய் உள்ளனர் என்பதே எதார்த்த உண்மை. 200 ரூபாய்க்குச் செல்பேசி (பழையது) கிடைக்கிறது. வாழ்நாள் 'சிம்' அட்டை இலவசமாகக் கிடைக்கிறது. நூற்றுக்கு நூறுபேர் செல்பேசி பயன்படுத்தும் காலம் அதிக தொலைவில் இல்லை.

செல்பேசியில் தமிழில் பயனர் இடைமுகம்

தமிழ்நாட்டில் செல்பேசியின் வளர்ச்சி இவ்வளவு இருந்தும், செல்பேசியில் மக்கள் தமிழைப் பயன்படுத்துவது அரிதாகவே உள்ளது. குறுஞ்செய்தி போன்றவற்றில் தமிழ் பயன்படுத்தப்பட்ட போதிலும், புழக்கத்தில் உள்ள பெரும்பாலான செல்பேசிகளில் தமிழ் இடைமுகம் கிடையாது. நோக்கியா போன்ற நிறுவனங்கள் தமிழில் பயனர் இடைமுகத்தை வழங்கியுள்ள போதிலும் மக்கள் இன்னும் பரவலாகப்

பயன்படுத்தத் தொடங்கவில்லை. பயன்பாட்டில் உள்ள தமிழ் இடைமுகங்கள் 'பயனர் தோழமை' (User Friendly) உடையதாய் இல்லை என்பதையும் இங்குக் குறிப்பிட வேண்டும்.

செல்பேசியில் தமிழ் இடைமுகத்துக்கான தேவை

மூன்று விழுக்காட்டுக்கும் குறைவான மக்களே கணிப்பொறி பயன்படுத்துகின்றனர். மேலும் ஓரளவு ஆங்கிலம் அறிந்தவர்களே கணிப்பொறியைப் பயன்படுத்துகின்றனர். எனவே கணிப்பொறியில் தமிழ் இடைமுகத்தின் (Tamil Interface) தேவை இன்னுங்கூட அதிகமாக உணரப்படவில்லை. கணிப்பொறியோடு ஒப்பிடுகையில் செல்பேசியில் தமிழ் இடைமுகத்தின் அவசிய, அவசரத் தேவை அதிகமாகவே உணரப்படுகிறது. அத்தேவைக்கான காரணங்கள் பல. அவற்றுள் மிகவும் முக்கியத்துவம் வாய்ந்த காரணங்கள் சிலவற்றைக் காண்போம்:

- (1) கிராம மக்களிடையே செல்பேசி மிகப் பரவலாகப் பயன்படுத்தப்படுகிறது. தமிழ் மட்டுமே தெரிந்த சாதாரண மக்களும் செல்பேசியை மிக அதிகமாகப் பயன்படுத்துகின்றனர்.
- (2) செல்பேசி குரல்வழித் தொடர்புக்கு மட்டுமின்றி, உரை வடிவிலான குறுஞ்செய்தித் தொடர்புகளுக்கும் மிக அதிகமாகப் பயன்படுகிறது.
- (3) செல்பேசி தகவல் தொடர்பு சாதனமாக மட்டுமின்றி, சிறந்த பொழுது போக்குச் சாதனமாகவும் பயன்படுகிறது. ஒளிப்படங்கள் (Photos), இசைப் பாடல்கள் (Music), நிகழ்படத் துணுக்குகள் (Video Clips), பண்பலை வானொலி (FM Radio), படப்பிடிப்பி (Camera) மற்றும் பல பயன்பாடுகளைக் கொண்டுள்ளது. மக்கள் எண்களை அழுத்தி, பிரியமானவர்களைத் தொடர்புகொண்டு பேசுவதோடு மட்டுமின்றி, பல்வேறு பட்டித் தேர்வுகளை (Menu Options) அழுத்தி மேற்கண்ட பயன்பாடுகளை இயக்கிப் பயன்பெறுகின்றனர்.
- (4) வருங்காலத்தில் செல்பேசி என்பது தகவல் தொடர்புக்கு மட்டுமின்றி சாதாரண மக்களின் பல்வேறு வகையான தகவல் தேவைகளுக்கும் பயன்படப் போகிறது. குறைந்த படிப்பறிவு கொண்ட கிராமத்து விவசாயிகள் வானிலை அறிய, விதை, உரம், பூச்சி மருந்து விலை அறிய, தானியங்களின் கொள்முதல் விவரங்கள் அறியச் செல்பேசியைப் பயன்படுத்தும் காலம் வெகுதொலைவில் இல்லை.
- (5) தற்போது செல்பேசியில் தமிழ் விசைத்தளத்தைத் (Tamil Keypad) தரப்படுத்தும் முயற்சிகள் நடைபெற்று வருகின்றன. அது நடைமுறைக்கு வரும்போது, செல்பேசியில் தமிழின் பயன்பாடு மேலும் பரவலாகும். அப்போது செல்பேசிக் கருவியில் தமிழ் இடைமுகத்தின் தேவை அதிகமாக உணரப்படும்.

தரப்படுத்தலுக்கான தேவை

பல்வேறு நிறுவனங்கள் தயாரிக்கும் செல்பேசிக் கருவிகளில் பல்வேறுபட்ட வசதிகள் இருக்கின்றன. ஆயிரம் ரூபாயிலிருந்து முப்பத்தையாயிரம் ரூபாய்வரை விலை கொண்ட செல்பேசிகள் பயன்பாட்டில் உள்ளன. தற்போது பயன்பாட்டில் உள்ள செல்பேசிகளை மூன்று பெரும் பிரிவுகளில் அடக்கலாம். வெறுமனே தொலைபேசி அழைப்புக்கும், குறுஞ்செய்தி அனுப்பவும் பயன்படும் செல்பேசிகள். இவற்றின் விலை ஆயிரம் ரூபாயிலிருந்து இரண்டாயிரம் ரூபாய்க்குள் அடக்கம். இரண்டாவது வகை ஒளிப்படம், இசை, பண்பலை வானொலி, படப்பிடிப்பி போன்ற வசதிகள் கொண்ட செல்பேசிகள். இவை மூவாயிரம் ரூபாயிலிருந்து ஐயாயிரம் ரூபாய்க்குள் கிடைக்கின்றன. 3ஜி, தொடுதிரை, வைஃபி, இணையம் போன்ற வசதிகள் கொண்ட உயர்நிலைச் செல்பேசிகள். இவை ஆறாயிரம் ரூபாயிலிருந்து பதினைந்தாயிரம் ரூபாய்வரை விலை கொண்டவை. 25 ஆயிரம் முதல் 35 ஆயிரம் விலைகொண்ட ஐஃபோன் போன்ற செல்பேசிகளும் உள்ளன.

இவ்வாறு பல்வேறுபட்ட செல்பேசிக் கருவிகள் பயன்பாட்டில் இருந்த போதிலும், அவற்றில் இடம்பெறும் பயனர் இடைமுகங்களில் அதிக வேறுபாடுகள் கிடையாது. இடைமுகங்களின் அமைப்புமுறையில் சிற்சில வேறுபாடுகள் இருந்த போதிலும் அவற்றில் காணப்படும் ஆங்கிலச் சொற்கள் பெரும்பாலும் ஒன்றாகவே உள்ளன. எடுத்துக்காட்டாக, Menu, Select, Options, Back, Exit, Cancel, On, Off, Yes, No, Switch off, Messages, Inbox, Outbox, Save, Send, Delete, Contacts, Search, Call, Missed Calls, Received Calls, Dialed numbers, Settings, Profile, General, Silent, Tones, Ringing tone, Ringing volume, Vibrating alert, Clock, Alarm போன்ற சொற்கள் அனைத்துவகைச் செல்பேசிகளிலும் அனைத்துவகை இடைமுகங்களிலும் காணப்படுகின்றன. எனவே தமிழ் இடைமுகங்களில் இவற்றுக்கு இணையான தமிழ்ச் சொற்கள் ஒன்றாகவே இருக்க வேண்டியது அவசியம். இடைமுகங்களில் தமிழ்க் கலைச்சொற்கள் வெவ்வேறு நிறுவனங்களின் கருவிகளில் வெவ்வேறு வகையாக அமையுமெனில் பயனாளர்களுக்குக் குழப்பமே மிஞ்சும்.

குறிப்பாகச் செல்பேசிக் கருவிகளில் தரப்படுத்தப்பட்ட கலைச்சொற்கள் பயன்படுத்த வேண்டிய தேவைக்கு இன்னொரு முக்கியமான காரணமும் உள்ளது. ஒரு செல்பேசிக் கருவியை அதிகப் பட்சமாக ஒருவர் மூன்று ஆண்டுகள் பயன்படுத்துகிறார். அவ்வாறு மூன்றாண்டு பயன்படுத்துவோரின் எண்ணிக்கை மிகக் குறைவு. இரண்டு ஆண்டுகளுக்குள்ளாகச் செல்பேசிக் கருவியை மாற்றிக் கொள்பவர்களே அதிகம். மூன்று மாதத்தில், ஆறு மாதத்தில் செல்பேசியை மாற்றிக் கொள்பவர்களும் பெருகிவிட்டனர். புதிய புதிய வசதிகளைக் கொண்ட செல்பேசிகள் நாள்தோறும் சந்தையில் அறிமுகமாவதும் அவற்றின் விலை நாளுக்கு நாள் குறைந்து கொண்டே போவதும் இதற்குக் காரணமாகும். செல்பேசிகள் தொலைந்து போவதும், களவு போவதும் மற்றொரு காரணமாகும். ஒருவரே ஒன்றுக்கு மேற்பட்ட செல்பேசிகள் வைத்திருப்பதையும் காண முடிகிறது. வாடிக்கையாளர்கள் வேறெந்தப் பயன்பாட்டுக் கருவிகளைக் காட்டிலும் செல்பேசிக் கருவியை அடிக்கடி மாற்றிக் கொள்கின்றனர். எனவே செல்பேசியின் பயனர் இடைமுகங்களில் காணப்படும் கலைச்சொற்கள் ஒன்றுபோல அமைய வேண்டியது அவசியமாகிறது.

ஆங்கில இடைமுகங்களில் கலைச்சொற்கள் அவ்வாறு ஒன்றுபோலவே அமைந்துள்ளன. அதுபோல தமிழ் இடைமுகங்களிலும் கலைச்சொற்கள் ஒன்றுபோல அமைய வேண்டும். எனவே செல்பேசிக் கருவிகளுக்கான தமிழ்க் கலைச்சொற்கள் தரப்படுத்தப்பட்டு, பொதுவான தமிழ் இடைமுகம் செல்பேசித் தயாரிப்பு நிறுவனங்களுக்குத் தரப்பட வேண்டும். அனைத்துச் செல்பேசித் தயாரிப்பு நிறுவனங்களும் தமிழ் இடைமுகங்களில் தரப்படுத்தப்பட்ட கலைச்சொற்களைப் பயன்படுத்திட முயற்சிகள் மேற்கொள்ளப்பட வேண்டும்.

தரப்படுத்தலுக்கான திட்டப்பணி

செல்பேசிகளில் தமிழ் விசைத்தளத்தைத் (Tamil Keypad) தரப்படுத்துவதற்கான முயற்சிகள் மேற்கொள்ளப்பட்டு வருகின்றன. தமிழ்நாடு அரசும் தமிழ் இணையப் பல்கலைக் கழகம் போன்ற நிறுவனங்களும் இம்முயற்சியில் ஈடுபட்டுள்ளன. இந்த மாநாட்டில் இதற்கான பரிந்துரை முன்வைக்கப்படுமென எதிர்பார்க்கப்படுகிறது. வருங்காலத்தில் செல்பேசிகளில் தரப்படுத்தப்பட்ட தமிழ் விசைத்தளம் அமையும் வாய்ப்பு உள்ளது. அதுபோலவே அமையவிருக்கும் தமிழ் இடைமுகங்களும் தரப்படுத்தப்பட்ட கலைச்சொற்களைக் கொண்டிருக்க வேண்டும். செல்பேசிகளில் தமிழ் விசைத்தளத்தைத் தரப்படுத்தும் திட்டப்பணியின் ஓர் அங்கமாகவே தமிழ் இடைமுகத்தைத் தரப்படுத்துவதற்கான முயற்சியையும் மேற்கொள்வது பொருத்தமாக இருக்கும். அதற்கான முயற்சியில் உத்தமம் போன்ற தமிழ் அமைப்புகளும், தமிழ்நாடு அரசும் முன்முயற்சி எடுக்கவேண்டும். இந்த முயற்சியை மேலெடுத்துச் செல்வதற்கான ஒரு குழுவை இந்த மாநாட்டிலேயே அமைக்க வேண்டியது அவசியத் தேவையாகும்.

End notes

¹ Information Note to the Press (Press Release No. 15/2010) by Telecom Regulatory Authority Of India

² Information Note to the Press (Press Release No. 15/2010) – Annexure-I by Telecom Regulatory Authority Of India

பின்னிணைப்பு

செல்பேசிகளில் தமிழ் இடைமுகத்துக்கான சில கலைச்சொற்கள்:

Menu	பட்டி	Voice Messages	குரல் செய்திகள்
Select	தேர்ந்தெடு	Picture messages	படச் செய்திகள்
Options	தேர்வுகள்	Info Messages	தகவல் செய்திகள்
Back	பின்னே	Service Commands	சேவை ஆணைகள்
Exit	நீங்கு	Delete Messages	செய்திகள் அழி
Cancel	விடு	Message Settings	செய்தி அமைவுகள்
On	நிகழ் / நிகழ்த்து	General Settings	பொது அமைவுகள்
Off	அகல் / அகற்று	Text Messages	உரைச் செய்திகள்
Automatic	தானே	Multimedia Messages	பல்லாடகச் செய்திகள்
Go to	அங்கு போ	E-mail Messages	மின்னஞ்சல் செய்திகள்
Names	பெயர்கள்	Contacts	தொடர்புகள்
Save	சேமி	Search	தேடு
Delete	அழி	Add new contact	புதிய தொடர்பு சேர்
Yes	ஆம்	Add new group	புதிய குழு சேர்
Ok	சரி	Edit contact	தொடர்பு திருத்து
No	இல்லை	Delete contact	தொடர்பு அழி
Help	உதவி	Move contact	தொடர்பு நகர்த்து
Show	காண்பி	Copy contact	தொடர்பு நகலெடு
Read	படி	Mark	குறியிடு
Switch off	அணை	Mark all	யாவும் குறியிடு
Messages	செய்திகள்	Unmark	குறியெடு
Create message	செய்தி எழுது	Log	பதிகை
Inbox	செய்திப் பெட்டி	Call log	அழைப்புப் பதிகை
E-mail mailbox	மின்னஞ்சல் பெட்டி	Call register	அழைப்புப் பதிவேடு
Drafts	வரைவுகள்	Recent calls	அண்மை அழைப்புகள்
Outbox	செல்மடல்	Missed calls	விடுபட்ட அழைப்புகள்
Sent Items	அனுப்பியவை	Received calls	பெற்ற அழைப்புகள்
Saved Items	சேமித்தவை	Dialled numbers	அழைத்த எண்கள்
Send	அனுப்பு	Message recipients	செய்தி பெறுவோர்
Dictionary	அகராதி	Clear log lists	பதிவுப் பட்டியல் அழி
Clear text	உரை அழி	Call duration	அழைப்பு காலம்
Save message	செய்தி சேமி	Message log	செய்திப் பதிவு
Exit Editor	தொகுப்பி நீங்கு	Settings	அமைவுகள்
Message counter	செய்தி எண்ணி	Tone settings	ஒலி அமைவுகள்

Delivery Reports	சேர்ப்பித்த அறிக்கை	Display settings	திரைக்காட்சி
Instant Messages	உடனடிச் செய்திகள்	Time settings	நேர அமைவுகள்
Call settings	அழைப்பு அமைவுகள்	Tones	ஒலிப்புகள்
Phone settings	பேசி அமைவுகள்	Recordings	பதிவுகள்
Security settings	பாதுகாப்பு அமைவுகள்	Media	ஊடகம்
Profile	வரைவாக்கம்	Camera	படப்பிடிப்பி
General	பொது	Radio	வானொலி
Silent	மவுனம்	Recorder	பதிப்பி
Discreet	ஏதேனும்	Organiser	ஒழுங்கமைப்பி
Loud	வெளிப்படட	Clock	மணிகாட்டி
Meeting	கூட்டம்	Alarm clock	எழுப்பு மணி
Outdoor	வெளிப்புறம்	Alarm time	எழுப்பு நேரம்
Theme	அழகாக்கம்	Alarm tone	எழுப்பு ஒலி
Tones	ஒலிகள்	Repeat alarm	திரும்ப எழுப்பு
Ringtone	மணி ஒலி	Speaking clock	பேசும் மணிகாட்டி
Ringtone volume	ஒலிப்பு அளவு	Calendar	நாட்காட்டி
Vibrating alert	அதிர்வு உணர்த்து	Notes	குறிப்புகள்
Msg.alert tone	செய்தி உணர்த்து ஒலி	Calculator	கணிப்பி
Next	அடுத்து	Timer	நேரங்காட்டி
Change	மாற்று	Stopwatch	நிறுத்து மணி
Date and time	தேதி நேரம்	Applications	பயன்பாடுகள்
Connectivity	இணைப்புநிலை	Games	விளையாட்டுகள்
Call	அழை	Collection	திரட்டு
Call divert	அழைப்பு திருப்பு	Web	வலை
Anykey answer	ஏதோவிசை பதில்	Home	முகப்பு
Automatic redial	தானே மறுஅழைப்பு	Bookmarks	நூற்குறி
Speed dialing	விரைவு அழைப்பு	Go to address	முகவரிக்குப் போ
Call waiting	அழைப்பு காத்திருப்பு	Download	பதிவிறக்கம்
Send my caller ID	என் அடையாளம் அனுப்பு	Language	மொழி
Phone	பேசி	Wallpaper	முகப்புப் படம்
Language settings	மொழி அமைவுகள்	Screen saver	திரைக்காப்பு
Automatic	தானியங்கு விசையரண்	Power saver	மின் சேமிப்பி
Security keyguard	பாதுகாப்பு விசையரண்	Charging	மின்னேற்றல்
Welcome note	வரவேற்புக் குறிப்பு	Reminders	நினைவுறுத்திகள்
Network selection	பிணையத் தேர்வு	Extras	உதிரிகள்
Start-up tone	தொடக்க ஒலி	Add new	புதிது சேர்
Gallery	கலைக்கூடம்	Unlock	திற
Images	படங்கள்	Converter	மாற்றி
Video clips	நிகழ்படங்கள்	Composer	இசைஅமைப்பி
Music files	இசைக் கோப்புகள்	Demo	வெள்ளோட்டம்

Software Architectures for Tamil Mobile Learning

S.Swarnalatha

24, K.G Gardens, Vartharajapuram,
Coimbatore-641015

E-mail: swarnaveni@yahoo.com

mobile no: 9600772878

Abstract

Our main objective in this part of the project has been to extend the distribution of Tamil language learning materials and communication to lighter equipment, specifically PDA and mobile phone. The challenge is then to develop the system and server side to present materials in ways suitable for PDA technology, find acceptable solutions for distribution of materials and for *administration to student, teacher to student/student to teacher and student to student communication*.

It is our aim in designing the environment for the mobile learner to extend and increase the flexibility of Tamil language education, that to some extent took a step backwards when converting from paper based to online learning, where students largely were required to study at a place (and at a time) where a computer with access the Internet was available.

Introduction

This paper examines what kinds of software architecture might be used to build Mlearning systems and outlines what factors and issues should be considered in terms of the benefits and drawbacks of each generic architecture. In the following section we review some relevant literature on key aspects of M-learning systems. We then outline a number of software architectures that may be used to build M-learning applications, looking at how each approach may contribute to the requirements of M-learning while considering the practical challenges and limitations. We provide a summary of the key issues associated with each architecture and suggest some recommendations for software architectures appropriate to different types of mobile learning system.

Non Adaptive Mark-Up

A number of mobile learning applications have been developed that use some specific form of browser mark-up for their client side presentation. This mark-up may be Wireless Markup Language (WML) or variations on the HyperText Markup Language (HTML) such as cHTML (compact HTML), XHTML (eXtensible HTML) Basic or XHTML Mobile Profile, depending on which types of mobile device are being supported. Regardless of the mark-up, the content may be served as static pages or generated dynamically on the server, using technologies such as server pages and/or eXtensible Stylesheet Language Transformations (XSLT) (Figure 2). Fowler classifies these to approaches to dynamic mark-up as the template (server page) and the transform (XSLT) approach, though in fact it is possible to combine the two, for example by using Tag libraries (such as the JSP standard tag library, the JSTL) in server pages that support transformations.

The advantage of this approach is that it is lightweight from the client device perspective requiring only the device's normal browser. The problem with non-adaptive mark-up is that using a particular mark-up language, for example WML, means that the content can only be rendered by browsers that understand that particular type of mark-up. Even though the content may be dynamically generated, it is only being generated for a specific type of client.

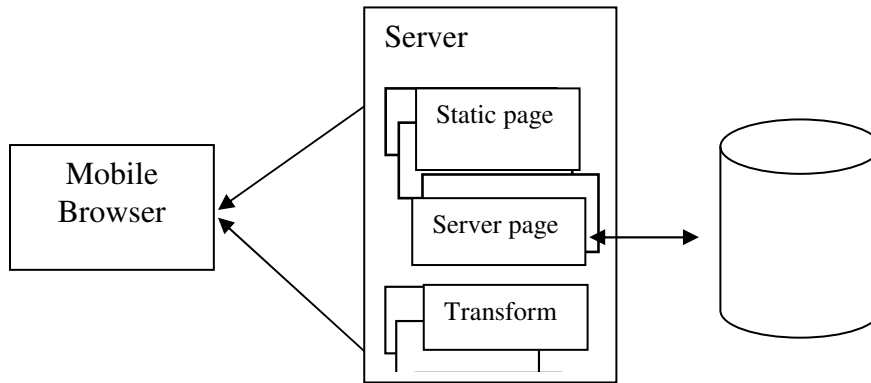


Figure 2: Non-adaptive mark-up architecture

Adaptive Mark-Up

Adaptive mark-up requires a server side process that is able to generate mark-up appropriate to the mobile device from a common set of contents. There are a number of approaches to this, for example an application can interrogate the HTTP (HyperText Transfer Protocol) header of the request and identify the client browser type from the 'user-agent' field, then generate client specific mark-up using various XSL transformations. An alternative approach is to use a tag library such as Wireless Abstraction Library (WALL), a JSP tag library that builds on the Wireless Universal Resource File (WURFL). WURFL is able to recognise user agent information and identify different devices, while WALL generates device specific mark-up. Either way, the advantage of this approach is that it enables an Mlearning application to support multiple types of mobile browser (Figure 3).

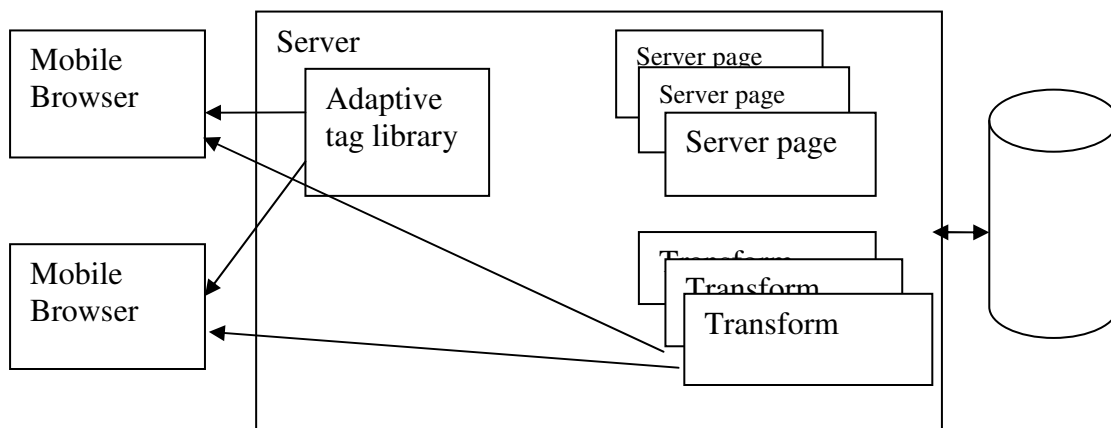


Figure 3: Adaptive mark-up architecture

Mobile Client Side Application

While the mark-up approach to M-learning architectures can provide a good range of content across a wide range of devices, confining the learner's activity to what can be supported by a mobile browser can limit the range of learning activities that are possible on the device. For example, interactive learning games may be difficult or impossible to support. An alternative approach to using mark-up to provide content via the mobile Internet is to provide applications that can be downloaded to the mobile device. There are three general categories of application that may be developed for mobile devices.

The first approach is an application written for a specific mobile device platform, for example targeting a particular model or make of mobile phone. These applications can take advantage of the special characteristics of that particular device, which can make them, for example, highly performant, but of course these applications cannot be used on other devices. A second approach is to use Microsoft Windows based applications, for example building an application using Windows Mobile components. This approach is more generic than writing for a particular device, since there are a number of devices that support Windows Mobile. However such applications cannot run on the majority of mobile devices, since there are many other operating systems being used including Palm OS, Symbian and Linux. The third, most generic approach, is to use Java Micro Edition (Java ME) . Using this software platform enables us to deliver a relatively rich client experience to a wide range of devices. Although Java ME has its limitations compared to some other application platforms such as Symbian and BREW, it works across a large proportion of mobile devices, including many Windows phones. In the context of mobile phones, the specific configuration and profile typically installed is the Connected Limited Device Configuration (CLDC) supporting the Mobile Information Device Profile (MIDP).

Java ME applications that run using this profile are known as MIDlets. Of course coding at the higher level of abstraction that enables interoperability has its costs in performance terms. For example Java ME applications have been shown to execute at about half the speed of equivalent C programs. There are also issues regarding the version of both the CLDC and MIDlet specifications that a given phone may support. Important differences between versions include floating point number support and security management, among many others. If a mobile client application is chosen as the software architecture, then the overall system design is simple.

The application runs standalone on the client, so the only role of the server is to enable download and installation of the application (Figure 4). This may be done either wirelessly or using a cable. Wireless download of Java applications can be provided using OTA (over the air) provisioning, a standardised approach that provides a consistent client-server interaction and enables version control for application downloads. The mobile application may use the data store on the mobile device but will not require access to any server side resources.

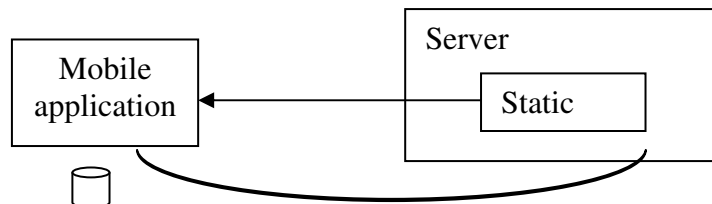


Figure 4: Client side application architecture

Smart Client with Server Connectivity

So far we have considered two very different types of software architectures for mobile learning systems, one based on providing a page based mobile Internet system, the other using a downloadable application client. However there is another approach that combines features from both of these architectures, the smart client that connects to the server. In this approach, there is a client side application, but that application does not run standalone. Rather, it communicates with the server to send and receive information while the application is running (Figure 5). There are a number of advantages to this approach. First, it is possible for the mobile learning application to provide a much wider set of content than is possible with a single downloaded application, since the storage size of the mobile device limits the amount of learning content that can be downloaded. A smart client that connects to the server can access learning content on demand without having to keep it all stored in the mobile device. Similarly, we can utilise the server to store information about the clients so we can, for example, maintain a sophisticated user profile on the server. In addition, the ability of the device to communicate data with the server in both directions (upload and download) means that it is possible to build a collaborative learning system where multiple users can send and receive data to and from each other.

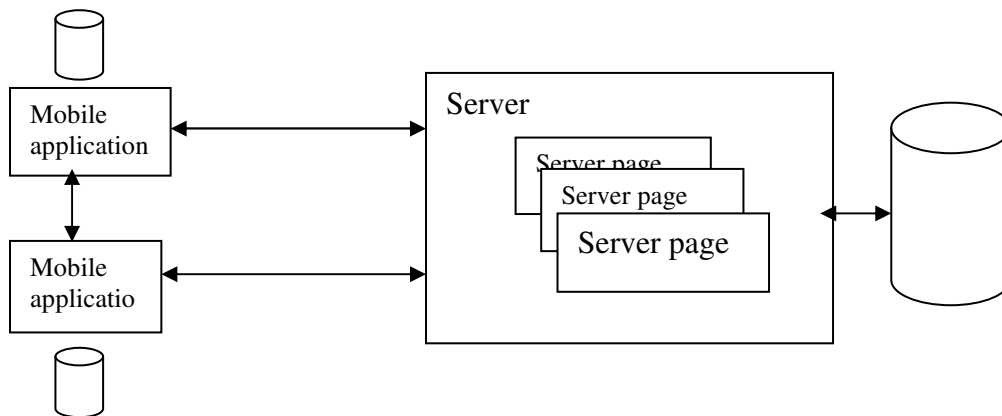


Figure 5: Smart client with server connectivity architecture

The price we must pay for this power and flexibility is complexity. For example, data management becomes a more complex issue because there may be data stored in a local data store on the client device as well as on the server. Mobile databases will need to be synchronized with central databases, and cache management has to be sophisticated to cope with the small memory size in most mobile devices. Distributed smart client applications using this architecture also need an appropriate communication protocol, such as XML over HTTP, to enable the client to access server side resources at run time.

Tamil Language Learning through Mobile

M-learning offers a powerful and practical solution to many learning and training challenges, such as:

- in collaborative projects and fieldwork
- as a classroom alternative to books or computers
- where learners are widely dispersed

- to engage with learners who in the past have felt excluded
- in promotional and awareness campaigns
- for 'just-in-time' employee training.

From a teaching and learning point of view, campus-wide internet access - or even access that targets social and learning spaces such as refectories, libraries, lecture rooms and labs - is what truly blends together online and face-to-face learning. It means that while they're on campus, a student can access their online learning just by turning on their net book or iPhone. They can contribute to class online discussions while eating lunch or access their readings before class, using the technology they already have with them: their laptop, net book, or other wi-fi capable mobile device.

For mobile language learning - and even for flexible learning - at any educational institution, equipping formal and informal learning spaces (such as social spaces) with fundamental enabling technologies like wireless internet access has to be at the top of the priority list. It even makes sense from a budget point of view, as every laptop a student brings in and uses takes pressure off the student labs. This, in turn, reduces the amount that has to be spent on standard-image, admin-locked, physical lab computers... and frees students to use their own computers which can be configured to best support their particular program of study.

Conclusion

In this paper we have provided a brief overview of four software architectures for mobile learning; non adaptive mark-up, adaptive mark-up, mobile client side application and smart client with server connectivity. All of these architectures have their own strengths and weaknesses, and in most cases we are trading flexibility against complexity. In addition there are different levels of server connectivity required for these different architectures, and successful applications depend not only on the technical infrastructure but also the social context within which issues such as pricing come into play. Therefore deciding on suitable software architecture for a specific Mlearning system depends not only on technical factors but also an analysis of the user context. While we may see that smart client architecture with server connectivity can provide us with the richest mobile learning environment, alternative architectures may prove easier to install, more robust in use, more easily deployed to a larger range of devices and cheaper for the learner to maintain. Therefore the most important aspect of designing an M-learning system architecture is to consider all aspects of the user context rather than just focus on the technical platform.

References

The use of palmtop computers for learning A review of the literature - Learning and Skills development agency.

The use of Computer and Video games for learning - Alice Mitchell & carol Savill-Smith.

<http://www.xenglobaltech.com/j2me-mobile-applications.html>

<http://www.nextwavemultimedia.com/html/mobilegaming.html>

http://learning.ericsson.net/mlearning2/project_one/wap_article.html

Software architectures for mobile learning D. PARSONS & H. RYU

Chang, C. and J. Sheu (2002). Design and implementation of ad hoc classroom and eschoolbag systems for ubiquitous learning. IEEE Int. Workshop Wireless and Mobile Technologies in Education.

Chen, Y., T. Kao, et al. (2002). A mobile scaffolding-aid-based bird-watching learning systems. IEEE Int. Workshop Wireless and Mobile Technologies in Education.

Coulton, P., O. Rashid, et al. (2005). "Creating Entertainment Applications for Cellular Phones." ACM Computers in Entertainment 3(3).

Domer, J., M. Nanja, et al. (2004). Comparative Performance Analysis of Mobile Runtimes on Intel XScale® Technology. 2004 workshop on Interpreters, Virtual Machines and Emulators (IVME'04), Washington, D.C., USA, ACM Press.

Predictive Tamil Short Messages for Handheld Devices

Abirami.S ,Vilvanathan. K and Baskaran. R

Dept of Computer Science & Engg, CEG, Anna University, Chennai – 25.

Email: abirami_mr@yahoo.com, vilva.mca@gmail.com, baaski@annauniv.edu

Abstract

This paper aims at providing ease and flexibility to Tamil SMS users by suggesting the commonly used message templates with few key strokes, based on the context of letter typed. This system intends to predict short messages in Tamil, as similar to T9 dictionary, but the scope is not restricted to the prediction of a single word. Rather this attempt to predict the short messages (sentences) communicated in our day-to-day life.

Initially, keypad of the mobile phone has been configured with Tamil Vowels and Consonants using Tamil keypad layout standardized for hand held devices. To start with, this predictive SMS system allows the user to type an initial letter in Tamil to predict the possible nouns and verbs starting with the letter typed. Based on the user selection, initial word gets composed. If the initial word appears to be a noun, probable actions (i.e. Verbs) associated with the noun with proper case endings could be predicted. Depending on the needs of the user, verbal actions could be enriched with suitable adverbs too. If the initial word appears to be an adverb, suitable verb could be appended with it.

Possible templates (nouns, verbs and adverbs) are stored in Unicode text format in a file (records). Depending on the context (taking nouns, case markers and adverbs into account), suitable records would be retrieved. Few English message templates are available in standard mobile phones to reduce the typing hindrance involved in messaging. Moreover, these templates are not applicable to Tamil language. As a result, this paper has been motivated to introduce a Predictive SMS system which intends to reduce the number of key presses required to frame a message in Tamil. In addition, this system covers most of the commonly used Tamil sentences and not restricted to fewer templates, thereby attempting to provide a full fledged predictive Tamil short message service.

Introduction

Now-a-days, SMS through mobile phones have become an integral part of human communication. To make it easier and innovative, it is important to have the facility to enter, send, receive and read SMS's in Tamil. SMS in Tamil are already a reality. This paper is therefore, we believe, the first attempt to come out with a possibility to send SMS's in Tamil, as comfortably as one sends an SMS in English. To add flavor to it, we try to make it more easily than English.

Tamil language

Tamil is a South Indian language spoken widely in Tamil Nadu in India. Tamil has the longest unbroken literary tradition amongst the Dravidian languages. The earliest available text is the Tolkaapiyam, a work describing the language of the classical period. There are several other famous works in Tamil like Kambar Ramayanam and Silapathigaram but few supports in Tamil which speaks about the greatness of the language. For example the Thirukural is translated into most other languages due to its richness in content. It is a collection of two sentence poems efficiently conveying and few other things in a hidden language called Slaydai in Tamil. Tamil has 12 vowels and 18 consonants. These are combined with each other to yield 216 composite characters and 1 special character (aayatha ezhuthu) counting to a total of $(12+18+216+1)$ 247 characters.

Vowels

Vowels in Tamil are otherwise called UyirEzhuthu and are of two types short (Kuril) and long(Nedil).

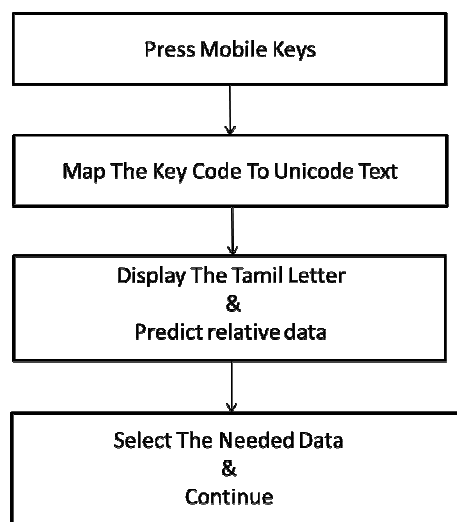
Consonants

Consonants are classified into three classes with 6 in each class and are called Vallinam, Idaiyinam, and Mellinam.

Tamil Unicode

The Unicode Standard (<http://www.unicode.org>) is the Universal Character encoding scheme for written characters and text. It defines the uniform way of encoding multilingual text that enables the exchange of text data internationally and creates the foundation of global software. The Tamil Unicode range is U+0B80 to U+0BFF. The Unicode characters are comprised of 2 bytes in nature. For example, the Unicode for the character **அ** is 0B85; the Unicode for the character **ஊ** is 0BAE+0BC0. The Unicode is designed for various other Tamil characters.

Functional Block Diagram



Taking Key Codes As Input

When the user presses the keys of the mobile phone, the key press event generates unique key codes for each key. Using those keys codes and the time interval between key press events, the appropriate Tamil letter is mapped and displayed on the screen.

Keyboard Mapping

Mobile Phone Keyboards have a limited number of keys (roughly 12). However, the use of the mobile phone as a messaging instrument (using SMS) has risen in the recent past. Currently, predictive texting schemes exist for English and many European and other languages. However, such a scheme does not exist for most Indian languages, including Tamil.

Moreover, the mobile phone keyboard for the English Language is generally arranged in the alphabetical order, that is the key for the number 2 contains the letters A,B,C; that for number 3 contains D, E, F; and so on. Following a similar arrangement is a relatively inefficient

scheme for Tamil; hence, we have come out with a more efficient arrangement that does not club frequently used letters on the same key. Our mapping assigns, in general, two vowel and two consonants to each key.

While pressing a key, the first letter assigned to that key is mapped. If the same key is pressed again, the time between the subsequent key strokes are calculated. If the time is below the quantum time assigned, instead of the first letter, the second letter assigned to that key is mapped. The same concept is applied for all the letters assigned to the keys.

When a consonant letter is pressed and subsequently if a vowel is keyed in, its corresponding consonant letter is mapped. For example, ma+e=mi.

Stored Records

The Possible Sentences used often are stored in the records, as far now, it is limited so as to make the work easier. The sentences are stored in Unicode format, thus making the size of records minimal.

Predict

As the keys are pressed and the corresponding Tamil characters are displayed in the screen, the system searches for the matching sentences in the Records and if any matches, they are displayed as choices. When the first letter is typed, the system searches for sentences starting with that letter in the record. And displays the possible sentences, the user can select one from that choice or can go for a different choice of typing in the second letter. Now the records are searched for sentences starting with that two letters combined. And the possible sentences are displayed. Now the user can select one from that choice or can go on for typing in the next letter. The same procedure follows for every letter typed.

Future work

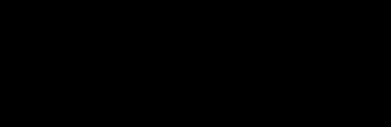
This paper limits the sentences in Records to a meaningful limit, we can increase the records in future for a better system. The user interface is also restricted to our convenience, it can be changed to a standard format.

Conclusion

This system encompasses the Tamil typing and prediction of the sentences as and when typing. This eliminates the typing and time to a considerable extent. Also it helps a person with less knowledge in Tamil typing and spelling.

16

தமிழ் ஒருங்குறி



Status of UNICODE

in The Indian Tamil Publishing Industry

P.Chellappan
Partner, Palaniappa Bros
chellappan@vsnl.com

The Publishing industry in general including the Tamil Publishing industry is at a crucial juncture today. The world of publishing is moving away slowly but surely from the traditional paper based publication towards e-book publication. This movement is more significant in English books than Tamil books. Nevertheless the Tamil publishing industry has to gear itself as soon as possible, so as not to fall behind and get lost in the technological revolution. The Tamil Publishing industry has traditionally been using 8 bit encodings just like all the other Indian languages. But looking at the future, it has to move over to 16 bit encodings.

The purpose of this study is to ascertain the current practices in the Tamil publishing industry so that suitable steps can be taken to help in this transition.

Methodology

A questionnaire was prepared and distributed to some of the major Tamil books, and newspaper publishers. The questionnaire mainly relates to the pre-press activities such as receipt of manuscripts, data entry, page layout, and graphic designing. The feedback was then analyzed.

Summary of analysis

In all 30 Book Publishers and 10 Magazine & Newspaper publishers were asked for their feedback. But only 11 responses were received from Book Publishers and 9 were received from Magazine & Newspaper publishers. It can be safely assumed that there will not be a significant difference in the findings even if the remaining feed back was received.

The data was analyzed and statistics are presented in the tables given below. In these tables (B P) refers to Book Publishers and (M & N) refers to Magazines and Newspapers.

Receipt of Manuscripts :

	Hard Copy	Soft Copy
B P	84.00%	16.00%
M & N	4.00%	96.00%

	TAM	TSCII	UNICODE	Others
B P	44.00%	12.00%	19.00%	25.00%
M & N	67.00%	-	-	33.00%

It can clearly be seen that in the BP segment manuscripts are received mostly as either hand written or typed documents. But in the M&N segment it is quite the opposite. 96% of the manuscripts are received as soft copies. These soft copies are again received in various font encodings. Fortunately, in most cases a single manuscript comes in a single encoding. So encoding conversion operations are reduced to a minimum, unless of course it is absolutely essential like in the case of manuscripts received in Unicode.

Type Setting :

	In-house	Outsourced
B P	23.00%	77.00%
M & N	100.00%	0.00%

	TAM	TSCII	UNICODE	Others
B P	44.00%	12.00%	19.00%	25.00%
M & N	67.00%	-	-	33.00%

	Page Maker	Indesign	Quark	Others
B P	57.00%	10.00%	25.00%	25.00%
M & N	10.00%	20.00%	10.00%	60.00%

Type Setting operation is one of the first major operations done by any publisher. Most of them have only limited in-house capability and the rest is outsourced. As can be seen in the tables above, the BP segment has 23% in-house capability but the M&N segment has 100% in-house capability. The major advantage of having in-house capability is in the tight control over font usage. In the case of outsourced operations, the publishers have little control over the font encoding used. Some times they are even forced to accept documents that use different font encoding since the DTP operator has only that option.

Some of the publishers are either ignorant or do not even bother about the font encodings, as they archive the books only as hard copies.

Page Layout :

	In-house	Outsourced
B P	27.00%	73.00%
M & N	100.00%	0.00%

	TAM	TSCII	UNICODE	Others
B P	50.00%	14.00%	-	36.00%
M & N	67.00%	-	-	33.00%

	Page Maker	Indesign	Quark	Others
B P	52.00%	23.00%	10.00%	15.00%
M & N	11.00%	78.00%	11.00%	-

Page Layout is an important aspect in the publication of any book, magazine or newspaper. It can be seen that in the BP segment 73% of this operation is outsourced. The reason is that they would like to leave it to the professionals. However in the M&N segment 100% of the page layout is done in-house. Newspapers and magazines have very little time on hand before the matter goes to print. Hence it makes sense to do this important job in-house.

One important aspect here is that UNICODE is not used in this final stage. This is attributable to the applications used for page layout. Page Maker, Indesign and Quark do not support complex script and hence cannot be used for Tamil, or for that matter any Indian language.

Graphic Design :

	In-house	Outsourced
B P	34.00%	66.00%
M & N	100.00%	0.00%

	TAM	TSCII	UNICODE	Others
B P	54.00%	7.00%	-	39.00%
M & N	67.00%	-	-	33.00%

	Photoshop	Illustrator	CorelDraw	Others
B P	57.00%	10.00%	25.00%	25.00%
M & N	33.00%	33.00%	33.00%	-

All the observations for Page Layout are exactly applicable for Graphic Design also. Here also the applications used viz. Photoshop, Illustrator and CorelDraw do not support complex script rendering. Hence Unicode encoded text cannot be used in this step.

Conclusion

The survey clearly indicates that the entire publishing industry uses only 8 bit encoded fonts like TAM and TSCII, with TAM being used more widely than the other encodings.

It also indicates that even when the original manuscript is received in Unicode, it is being converted into one of the 8 bit encodings for further use. This is mainly due to lack of support of complex script rendering in the applications that are used in the publishing industry.

The lack of awareness of UNICODE amongst the book publishers is also another contributing factor for non-usage of UNICODE. The reason for this lack of awareness can also be attributed to the non-usability of this encoding.

However all the publishers in the M&N segment are aware of UNICODE and in fact about 55% of the publishers in the M&N segment use UNICODE for their on-line editions. They refrain from using UNICODE in the print editions only due to the lack of support for complex script rendering in the applications that are used by them.

In the coming days when use of multiple languages in a single publication becomes more common, it is better to migrate from the legacy 8 bit encodings to 16 bit encodings. The inhibiting factor as already stated above is the lack of support for complex script rendering. Since a definite time frame is not available for provision of this support, usage of an all character 16 bit encoding like TACE16 will help in immediate migration to the 16 bit environment. One additional advantage of TACE is that even off-the-shelf e-book readers will be able to render Tamil e-books embedded with a TACE font.

Problems of using Unicode in software components and in NLP

Dr V. Krishnamoorthy
(Former professor Of Anna University)
5, Srivatsa Apartments
A1-11, 23 rd Cross Street
Besant nagar, Chennai 600 090
Email: prof.vkrish@gmail.com

Abstract

To speed up software development and reduce cost, using specialised software components is an established practice. Tamil editing software developers have been using a component called “text control”, to include functionalities like justification and tables. These components do not support complex scripts. Hence they are not usable for using Unicode Tamil in the software. It is not known when such support will be provided. Economic considerations may indicate that it may not come through in the immediate future.

The processing of letters, words, sentences, paragraphs and passages to ‘understand’ what they mean, and, produce answers to the required questions is the basic aim of Natural Language Processing (NLP). This has string processing as the core. The encoding of Tamil in Unicode is not very conducive to this string processing is shown by taking a few simple fundamental operations. The complex behavior is the result of the variable length encoding and unnatural way of representation of letters in Unicode.

1. Introduction

It is now widely known that Unicode Tamil is not supported in high-end publishing software. But less known is the fact that some software components also do not support Unicode Tamil. This may pose restrictions on developing specialised Tamil software in the long run. Here we highlight this problem by giving a specific example. We find that the reason for not supporting Unicode Tamil is that Tamil in Unicode needs level 2 implementation.

There are many special editing software for Tamil, which provide many other special features not available in Word etc. All these software use software components called “Text control”. This text control component is essential to provide the functionalities like, justification and tables. It is not possible to create such components on our own, for our own use, in view of the complexity involved. Tamil software developers, like many other developers world wide, use these components, to provide rich functionalities within their software.

When Unicode is used in these components, the following problems arise. Complex script support is not available in text controls, which are available today. Absence of this leads to incorrect alignment and undesirable and unacceptable editing experience. As such these components cannot be used for Tamil

Unicode. The component developers are not willing to specify when they will provide support for complex scripts. It is seen that these components work with the Tamil All Character Encoding (TACE).

In the second part, the differences in dealing with the current Unicode and the all character encoding, in the context of natural language processing is brought out. Specifically how the fundamental operations are done easily and quickly in the all character encoding is studied. This shows that keeping the data in the all character encoding will speed up the NLP in Tamil.

2. Using Tamil Unicode in components

Software components called Text Controls are used in many Bilingual text editors. Currently they do not provide full support for Tamil. They open the file with Tamil Unicode characters. But the visible cursor position does not correspond to the “actual” cursor position. If one tries to insert a letter or string at a particular position, especially near the end of a line, it gets inserted somewhere else. An example is shown in Figure 1.

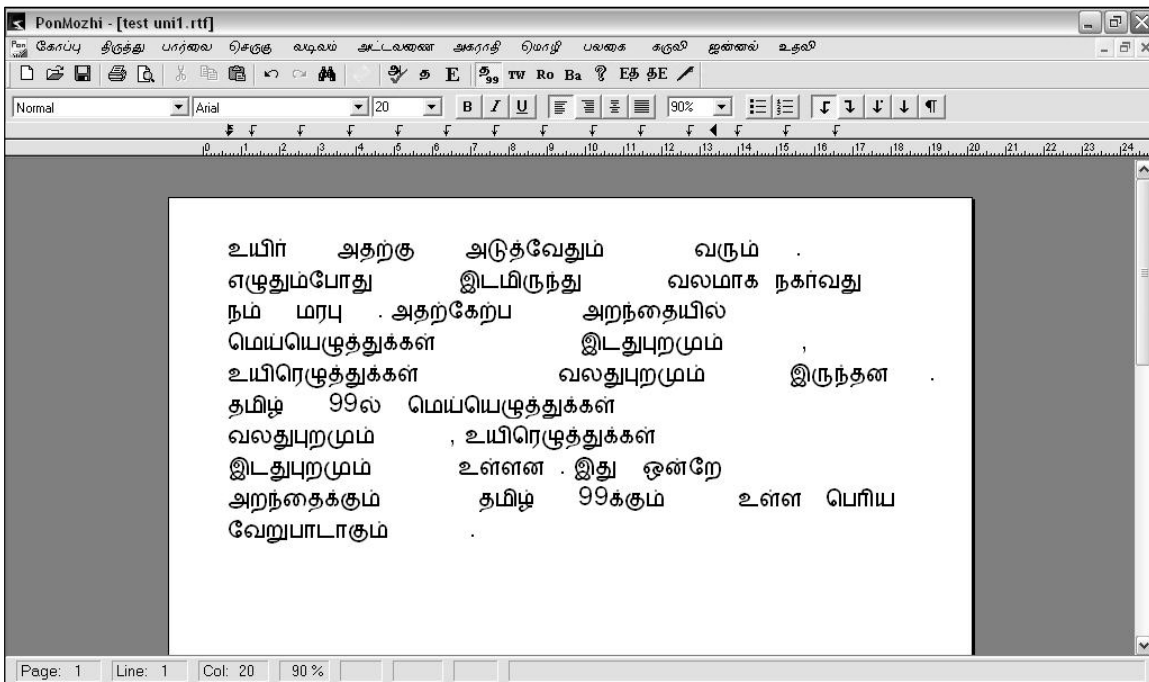


Figure 1

Here the cursor is placed at the end of the third word in the first line, which is ‘aduththum’, and the letter ‘vee’ is inserted using the ‘paste’ operation. The letter gets inserted in the middle of the third word.

It seems that this is due to the wrong calculation of the width of the letters. Here it seems that the letters are treated as having constant width. This can be seen by keeping the cursor at the left and tapping the right arrow key repeatedly. Also if ‘select all’ is given, the selected portion is shown wrongly. See Figure 2 for the left aligned text. The mismatch is clearly visible. Similar mismatch is seen in the case of the justified text also.

It is obvious that no commercial software can be sold with such grave shortcomings.

When contacted, the Text Control company has not given any definite answer about providing support to Tamil. Software components are not sold like mass products. Only software developers buy these. So the number of buyers will not be very large. When it comes to a particular language, this number may be very small. This may discourage many of these component vendors, who are not big companies, from providing support for a complex script, since it may involve considerable cost.

This indicates that no Tamil application using Tamil Unicode can make use of many good components in the near future. Whereas most of them may work without any difficulty with the all character encoding.

3. Problems in NLP

Comparing two strings is a fundamental operation, used heavily in NLP. For example, when a new word has to be added to an existing dictionary, the new word has to be put in the proper place. Otherwise it will lead to linear search, which is time consuming. When we compare two strings, it is natural to expect that the natural ordering of letters in Tamil be followed. This is essential, as the sorted sequence of words may be required to be printed for the human perusal, or stored for use in some other application.

Comparing two Tamil strings in Unicode is a formidable task, since the ordering of characters in Unicode does not follow the Tamil ordering. Also, the letters ksha and sri, which we consider as single letters, are stored as a combination of two or more characters. In the all character encoding, this task is just as easy as saying 1,2,3. The pseudo code for this can be given below.

```
Let minLength = minimum ( length(string1), length(string2) )
```

```
For I = 0 to minLength-1
```

```
    If (string1[I] > string2[I]) then string1 is bigger. Exit
```

```
// here the common part is same.
```

```
If ( length(string1) > length(string2) )
```

```
    then string1 is bigger.
```

```
    else string2 is bigger.
```

```
Exit.
```

Writing the pseudo code in the case of Unicode is not so easy. I do not attempt it here, due to lack of patience on my part. But, note that the following things have to be taken care of while comparing.

The consonants with 'a', have to be rearranged according to Tamil sorting order.

The mei pulli should come before any of the vowel mathras.

A consonant with 'a', with the pulli coming next to it, precedes that consonant with 'a'.

Ksha series letters should not be put in the ka series, though the first character is ka.

Sri has to be put in a separate place.

The length of the string is not the same as the number of letters. String length cannot be used straightaway.

One may understand the complexity and the time involved in writing this code. The time involved in executing these two codes is obvious.

Let us take another simple example.

Hyphenation is a problem faced in the printing industry. Let us consider a simple solution for this, with the following rules. Break only when there are 4 or more letters. At least 2 letters should be on both sides of the cutting point. The second part should not start with a pure consonant. (This solution has been implemented by us already.)

In the case of all character encoding, the pseudo code would look like this.

```
If ( length ( word ) < 4 ) exit.
```

```
// length is 4 or more, here
```

```
// try to accommodate as much as possible in the first line
```

```
for I = length(word) -2 to 2 do
```

```
    if word[I+1] is not a consonant, and, first I letters fit in the first line,
```

```
        then split after the I-th letter. Exit.
```

Considering the same solution in Unicode, the following things need attention.

The length of the string does not give the number of letters. Hence the condition that at least two letters on both sides has to be taken care of separately. The boundary points of the 'for loop' does not ensure this.

The cutting may happen at the middle of the letter. So every time it should be assured that we do not cut at the middle of a letter. Special attention is needed for sri and ksha series letters.

All these will need more coding and more execution time.

We will close with one more example.

Searching for a string as it is, has limited scope in Tamil. In Tamil, words appear with many variations. For example the word 'maram' appears as 'maraththil', 'maramaaka' etc. To find out all the places in which 'maram' appears, it has to locate all the words with variations also. A simple solution will be to first search for the string 'mara', and then find out whether the word found is a required word. For this start checking whether the next letter is 'im' or 'th'.

The same logic works for both Unicode and TACE. But the difference is the number of false hits. In Unicode, when 'mara' is searched, all the words with 'mar', 'mara', 'maraa', 'mari', ... will be the results of the search. This will result in a large set which has to be filtered for the required words. But in the case of TACE, only words having 'mara' will be found out. Others will not appear in the searched words. This will save the time for processing the unwanted words to a great extent.

In the three simple examples provided, we have shown that in the case of Unicode, the coding may be complicated, and even when the coding may be the same, some unnecessary work may be done. In view of the above, we can conclude that the processing for NLP will be easier and faster, and the running time will be less when TACE is used, when compared with Unicode. Also the internal data, like dictionary, kept in TACE will make the processing easier and quicker.

Conclusion

It is shown that some useful software components, which are being used by Tamil editing software developers, do not support Tamil Unicode. The complexity of executing fundamental operations of NLP using Unicode is brought out. These are due to the level 2 implementation of Tamil in Unicode, and, the variable and unnatural way the encoding is done. It is shown that the all character encoding provides a simple natural way of encoding, and hence provides better results.

References

1. Unicode.org
2. www.tamilvu.org/coresite/download/TACE16_Report_English.pdf

Challenges of Publishing Industry and E-Governance with Tamil Unicode (TU) and possible remedies with Tamil All Character Encoding in 16-bit (TACE16)

A.Elangovan

email : elango_ae@yahoo.co.uk,

Chair-Infitt-WG08 & Founder President, Kani Thamizh Sangam

MD-Cadgraf Digitals & Digiscape Gallery, Chennai, India.

Back ground - 21 years experience in Publishing and Printing, development of Multi-lingual Editorial Work flow system, Indian language fonts and interfaces for Win & Mac, Adobe plug-in development, Tamil hyphenation &, spell checking

Introduction

India is emerging as the Publishing hub of the world with 45% of publications in English, the world's third largest English publishing country after UK and USA. The remaining 55% is published in Indian languages. Today Print media is still the largest media in terms of readership, circulation, published pages and editorial manpower involved. Today publishing does not stop with print media but also extends to web, mobile, palm readers and broad-cast media.

E-Governance in India is in its nascent stage. Indian language enabling is essential for the successful implementation of any E-Governance project. E-Governance will include all media including web, print, mobile and palm readers. The data should be viewable, portable and reliable across various media, operating systems and applications.

Multilingual publishing is still a major challenge for publishers in spite of the Unicode advantage of having all languages of the world under one unified encoding scheme.

Today most of the content available in the web is in Unicode and Tamil Unicode is widely used by all segments of people. In spite of its popularity, the Publishing Industry and E-Governance segment face several challenges in implementing Tamil Unicode. This paper explains the specific areas of challenges and the possible remedial measures to some of them. It also forewarns the users about the pit falls in the implementation of Tamil Unicode and how to avoid them.

The paper discusses some of the implementation issues faced in real life projects with Tamil Unicode and the remedial measures taken. The paper also covers some possible solutions which include Innovative methods of Unicode Font design, TACE16 based applications and Software Plug-in Tools and its benefits.

Challenges in Publishing with Tamil Unicode

1. Text Editing

Today the editors have a wide choice of sources for the news, stories, events and statistical data. The sources include wire services, news agencies, reporters, journalists including freelancers, web sites, blogs, TV channels, digital archives and printed archives. The text data exchange mostly happens in the digital form through email and in other formats like pure text, word, spread sheet and PDF. More and more content in Tamil Unicode (TU) is available in the web, email and word formats and is becoming a popular source of information. However the text need to be converted to 8-bit legacy encoding like TAB/TAM before it can be paginated as most of the publishing applications like Adobe, Quark and Corel do not support TU yet. Converters are widely available to convert TU in pure text format and the availability for other formats like word is limited. We face two problems while doing this conversion. First, if the TU text contains more than one language ie. Tamil and English, the text in English gets converted to junk. This requires manual intervention to identify the English text, apply the English font, proof read the text and correct all junk characters before they are sent to pagination. The second one being loss of all formatting done in other applications like word while doing this conversion. The typical text process flow is given in (Fig.1).

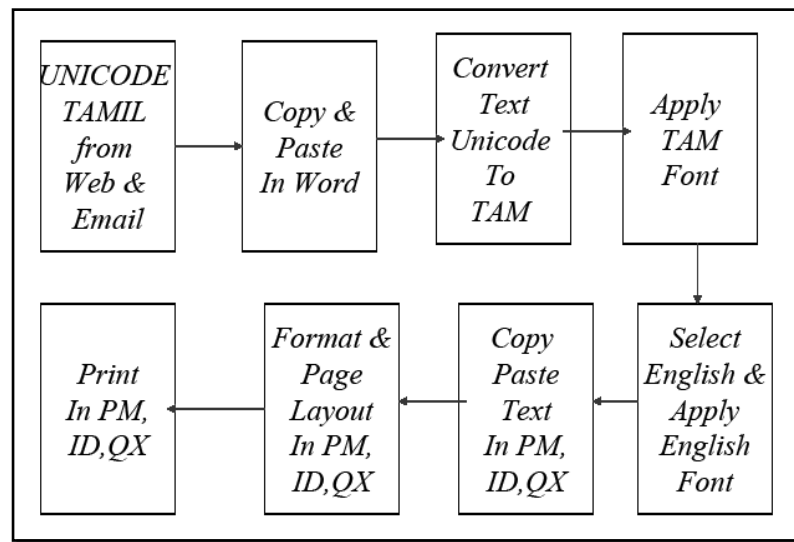


Figure 1 - TU Text Editing Process Flow

Solutions : One way to overcome the problem of multi-lingual text is to convert Unicode text to TACE16 (Tamil All Character Encoding for 16 bit) instead of 8-bit TAB/TAM. TACE16 can easily co-exist with other language text in Unicode. It is possible to retain the formatting features of the source document with help of plug-ins for conversion from within the respective applications. Eg. Indesign Encoding Converter plug-in.

2. Pagination

MS Word and Publisher are used to some extent in the entry level and do support Tamil Unicode with some limitations. PageMaker has been discontinued by Adobe few years back and anyway it does not support 16-bit encoding like Unicode. The most widely used pagination applications by

professional publishers like Newspapers, Magazines, Books, Directories, Government publications like budget documents, gazette publications, assembly documents are Adobe Indesign, Quark Xpress and Corel Draw. Unfortunately all these professional applications which are vital for publishing and printing are yet to provide TU support even in their latest versions (Fig.2 A). And the definite date of their future support is also not yet known. Also about 30% of the publishing industry uses Apple Macintosh systems, where the TU support is incomplete.

Solutions : Interestingly TACE16, which is the standard recommended by TN Task Force is found to be working well in all the above professional applications (Fig 2 B). It is also supported both in Windows and MacOSX operating systems. Real life trials in production have proved not only its support but also its higher efficiency.

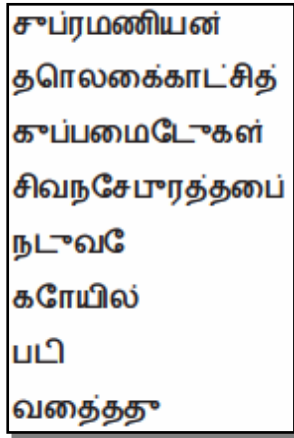


Figure 2 A - TU in Indesign & Quark

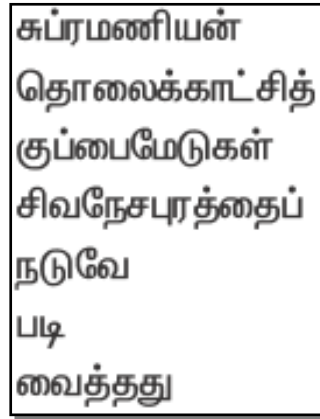


Figure 2 B - TACE16 in Indesign and Quark

3. Illustrations

Graphic illustrations like logos, graphs, charts, cartoons, sketches, back grounds, banners and book illustrations form important elements of publishing for not only print but also for web and mobile. The popular applications used for creating graphic illustrations are Adobe Illustrator and Corel Draw. Unfortunately both of them are yet to provide TU support. (Fig.3 A and 3 B)



Figure 3 A - TU in Corel Draw



Figure 3 B - TU in Adobe Illustrator



Figure 3 C - TACE16 in
CorelDraw and Illustrator

Solutions : Fortunately TACE16 is found to be working well in both these applications in both Windows and MacOSX. (Fig.3 C)

4. Image Editing

Photo editing, painting, colour correction, titling and labeling images are the next important functions of publishing. The leading application for image editing for print, web and animation is Adobe Photoshop. Corel Photo Paint is also used to some extent. Both these packages are yet to provide TU support.

Solutions: Once again TACE16 is found to be working well in both these applications in both Windows and MacOSX.

5. PDF Data Storage

Adobe PDF, the most popular Portable Document Format is almost the de-facto standard in print and publishing for creating the print ready format. PDF provides the portability of printing the documents across any printing device irrespective of the application and operating system in which it was originally created. It provides option for embedding the original fonts making it possible to print in any device without the need for having the original fonts at the printing end. PDF made it possible to electronically transfer print ready document across the world, paving the way for phenomenal growth of the growth of the e-Publishing (Electronic Publishing) industry. PDF is also the most popular format for storage of documents both for short term and long term. PDF retains the original text and graphics, which can be extracted and edited for future use. ISO standards for PDF are well defined and extensively used by the print and publishing industry while defining specifications and printing standards. Adobe PDF does not provide support for the complex rendering process in Tamil Unicode. The Unicode text is stored inside PDF in its own native format, different from that of the Unicode rendering order. When the Tamil Unicode text is extracted from the stored PDF documents, and placed back into MS Word, some characters with Egara, Eegera, Ugara, Uugara and Augara modifiers get reordered and sometimes even joins with the previous consonants, making the text meaningless (Fig.4 A).

Unicode PDF Text in Word
சுப்ரமணியன்
தொலைக்காட்சித்
குப்பைமேடுகள்
சிவேநசபுரத்தைப்
நடுவே
கோயில்
படி
வைத்தது

சுப்ரமணியன்
தொலைக்காட்சித்
குப்பைமேடுகள்
சிவேநசபுரத்தைப்
நடுவே
கோயில்
படி
வைத்தது

Figure 4 A - TU PDF Extracted in Word

Figure 4 A - TACE16 PDF Extracted in Word

Hence Tamil Unicode text stored in PDF standard is unreliable. The data integrity is totally lost in a round trip PDF storage and retrieval cycle even from MS Word. This is a major draw back for not

only print and publishing but also for all office and commercial applications where PDF is used as data storage and exchange format.

Solutions : Unlike Tamil Unicode, TACE16 stores all data as *Simple Script* without the use of the modifiers. All characters are stored as a single unit and there is no need for the OS to do the Glyph reordering process. Tamil text in TACE16 when stored in PDF, the internal storage order of PDF matches with the TACE16 ordering. On extraction of TACE16 from stored PDF documents, the data integrity is strictly maintained. TACE16 is found to be very reliable for round trip PDF storage and retrieval cycle. Production trials in real time publishing environment also proved its reliability with all applications like MS Word, Adobe, Quark and Corel in both Windows and MacOSX. (Fig 4 B)

6. Tamil Spell Checking

Automatic spell checking is a tool used extensively by not only the print and publishing but also by all office and commercial applications. Spell Checkers indicate not only spelling errors but also sandhi errors which is common in Tamil. Some of the text control tools used by the developers of Tamil Spell checkers do not support TU. Dual representation of certain characters in TU creates problems to the spell checking tools in identifying the words and its component parts. Most of the Spell checking engines use their own internal storage format for processing the text. This creates an additional burden in converting the Unicode text to internal encoding and after processing reconvert to Unicode format. This may affect the efficiency of the spell checking process to a great extent.

Solutions : TACE16 with all Tamil characters encoded with one to one mapping and without any complex rendering process is supported in all text control engines. In TACE16, there is no dual representation of any Tamil character. Hence TACE16 can be used as the internal encoding for processing also. The vowels and consonants are easily identified from their code values, which enhances the text processing speed considerably. Tamil Spell checking tools based on TACE16 not only work well with all current applications but also found to be far more efficient as compared to tools based on Tamil Unicode. Processing speed is very critical in production environments like newspapers.

7. Tamil Hyphenation and Justification :

Hyphenation and justification for Tamil text is far more complex than English. For English as the Hyphenation dictionaries are generally built into most of the professional applications, the hyphenation process is simple and automatic. For Tamil, there are no standard hyphenation dictionaries available. In a multi-column newspaper or magazine publication, hyphenation and justification is very critical. Improper word breaking at the end of each line will give very awkward look. Normally it is essential to do the manual hyphenation break at appropriate places. In real life production environment, reformatting and repagination are quite common. At each reformatting stage, the hyphenation breaking needs to be redone. All the earlier hyphenation breaks need to be removed as word breaks in middle of the lines is not allowed. With TU some of the complex characters can break in between the characters resulting in illegal breaks.

Solutions : Auto hyphenation for Tamil is normally achieved by plug-ins to applications like Indesign, etc; A series of hyphenation rules are built into the plug-ins, which automates this tedious

manual process. While applying these hyphenation rules, it is important to ensure that illegal breaks do not happen in between complex characters. With Tamil Unicode, the bit length varies from character to character and hence this process is little more complex. Whereas with TACE16, all characters are having uniform bit length, the process is much simpler. Also the possibility of illegal character break is totally eliminated in TACE16, as all characters are represented by a single code point.

8. Challenges in E-Governance with Tamil Unicode

Most of the e-governance projects require people involvement and interaction at grass-root level. Local language enabling in an efficient method is essential for the successful implementation of e-governance projects. An end to end e-governance project will involve multiple media which include web, print, mobile, palm readers, kiosks and e-books. The data should travel reliably across multiple operating systems and databases. The documents should be viewable, portable, text extractable for local processing and printable as and when required. The system should provide for reliable and efficient long term storage of data, which are independent of operating systems or applications. The system should provide a reliable and easy to use search and retrieval of the documents.

Tamil Unicode implementation in E-Governance is faced with a series of challenges. TU support is lacking in several media like print, mobile, ebooks, palm readers and broadcast. Wherever the documents need to be published in printed form, the documents need to be converted back to legacy 8-bit encoding. This increases the risk of errors and necessitates manual proof reading and correction. Integrating mobile, which is the most widely used device in the grass-root level, poses several limitations with TU due to lack of support for complex rendering in all devices uniformly.

Also reliable data portability across operating systems and applications is doubtful, as the TU support is incomplete in several applications. Dual representation of certain complex characters poses a problems to database sorting and indexing, resulting in inefficient search and retrieval of documents. Long term storage of portable documents with TU in PDF format is unreliable, as data integrity on retrieval can't be ensured. Also the stored data in text and other formats can't be retrieved without the help of the respective rendering engine (like *Uniscribe in MS*) provided by the respective operating system. This is a threat to data security in case of discontinuity of support for particular rendering engine by the OS provider after some years.

Solutions : TACE16 was tested by the National Informatic Center(NIC), GoI, as part of the testing process of TN Task Force on All Character Encoding in many of the E-Governance applications and confirmed its usability. Hence TACE16 could be used safely in all E-Governance applications where Unicode support is lacking or doubtful.

9. Font design for Multilingual Publishing

Unicode provides code points for all languages of the world including Indian languages. This give the provision for keeping multiple language scripts in a single font. (eg. Arial Unicode). One can design a font with both TU and TACE16 in the same font, which will give many operational advantages where both are required. This will help the user to read the TU text and convert to TACE16 when necessary for printing without the need for changing the font. Publishers of multi-lingual dictionary, commercial applications which require multi-lingual user interface and

government departments which need to create documents in multiple languages can create a font with the required language scripts to suit their special needs. This is one of the major advantages of Unicode as compared to legacy 8-bit encoding.

10. E-Paper Publishing

In today's condition Newspapers and magazine publishers who are using professional publishing applications will be forced to use TACE16 until TU support is available. Publishers who are publishing E-Paper version of their publication in the web (Electronic Paper in Newspaper look and feel) can use TACE16 encoding itself without the need to convert to TU. This gives the advantage of better access speed to the web readers. They also provide archived older newspapers for their readers. The extent of old archived E-papers one can provide depends on the online storage space provided in the Web Server. With TACE16, one can provide much longer period of newspaper archives online with available space.

11. Mobile Publishing:

Today's newspapers do not stop with print medium, but extend to mobile for quicker delivery of hot news and headlines. Tamil Unicode support in mobile devices across brands and models is still a distant reality. This is due to the difficulty in accommodating the overload of complex rendering necessary for TU. TACE16 being a simple script without any complex rendering, can be used easily and efficiently in mobile publishing.

12. E-Book Publishing

Book publishers may be using TACE16 due to lack of TU support in publishing applications. Just like in mobile devices, E-Book readers also lack the capability to handle the complex rendering required for TU. Here again TACE16 has proved its usefulness in field testing with few manufacturers.

13. Broadcast Media

Many media houses today are planning to integrate their newsroom for print with broadcast studios in order to provide cross media news coverage. They use special purpose titling and editing tools, which are yet to provide TU support. In these situations, TACE16 will be useful, as all these broadcast applications support simple scripts like TACE16.

Conclusion

Publishing and E-Governance, though faced with several challenges in implementing Tamil Unicode, they can be overcome to a great extent by use of the TACE16 encoding wherever TU support is lacking and by using suitable plug-ins for encoding conversion, spell checking and hyphenation tools.

A study on Tamil Script in digital media

N. Anbarasan

Chief Executive Officer

APPLESOFT

#39, 1st Floor, 1st Cross, 1st Main

Shivanagar, W. C. Road, Bangalore – 560010, INDIA

Tel : 23386167, Telefax : 23357167, Mobile : 9448053137

email - n.anbarasan@gmail.com, aplesoft@yahoo.com

Synopsis

Tamil script is one of the earliest recognized script amongst Indian scripts to get implemented on Computers, Pagers, Mobile phones, Dot Matrix Printers, Display boards etc and has been adopted for various applications like Desk Top Publishing, Messages, Web pages, Teaching/learning software, Billing software, Video sub-titling, News readers etc. This paper presents the possible basic requirement, which enabled implementation of Tamil script in digital media.

Whenever the ever growing technology has posed limitation in adopting Tamil script, its simplicity enabled implementation into such newer devices. But, in certain cases the number of glyphs required to implement Tamil script is definitely posing difficulty. This author analyses the difficulties posed by the Tamil script in implementing Tamil script on power hungry digital devices.

In order to implement Tamil script on digital devices, standards are required for input, storage and display. Eventhough, there are certain standards made available for Tamil script by concerned bodies, this author experienced contrasting nature of the standards prescribed for Tamil script. This author presents the experience gained while implementing the various standards as developer and as an user.

As a way out of possible seamless implementation of Tamil script and to have feedback for users while inputting Tamil text, the author suggests the possible script reform, the encoding standard and an input method. This paper also presents the possible fallout in adopting the script reforms and its possible implications on the existing implementations.

Introduction

Publishing being an earliest application, for which early hardware and software solutions were developed first to support Tamil script and still continue to be demanding. The next application identified was Word processing. It is evident that Publishing and Wordprocessing were driving the widespread use of computers for Tamil script related works. Also, some indigenous software were developed for publishing and word processing and some add-on software were also developed to enable to use the existing or familiar software developed for English. Over the years, as the capabilities of the Operating Systems were increasingly providing better facilities, software were developed using Tamil script for various requirements such as Teaching/Learning, Games, Video sub-titling, Billing, News readers etc.,.

Technologies behind the possibilities

The switchover from Vector graphics monitors to Pixel graphics monitors has opened revolutionary possibilities for the graphics based computer applications such as DTP, digital graphics, Video editing etc. The Pixel based graphics enabled development of various font technologies, which has thus resulted into soft fonts and then advanced from TTF to OTF. Apart from the Operating System based font technologies, the developers have also developed their own font technologies to provide support for Tamil script in their own software. Also, some developers provide the rendering engines to make the software, platform independent.

Basic requirement

In order to provide support for a script in any software, a well defined standard or atleast its well established implementation called industry standard is required. In the absence of a standard, the proprietary script implementations leads to incompatibility of the data crippling data exchange. In Operating Systems, which provide support for scripts by means of fonts, the scripts of languages intended to be used have to be encoded as glyphs in the fonts with the total number of glyphs for a script not exceeding the usable or available code positions of the code page. As the standards established to provide language support on computers are based on the scripts of the language, the standards are expected to provide well defined rules to provide feedback for various typing layouts and letter formation.

Standards prescribed for Tamil letters

Govt of India, in its efforts to promote usage of languages on computers have recommended standards for storage and keyboard layout for typing based on the recommendations of the committee constituted and these recommendations were announced as Bureau of Indian Standards, known as Indian Script Code for Information Interchange (ISCII). But, ISCII has never been implemented on popular Operating System such as Microsoft Windows series as an encoding for general use due to the difficulties associated with its implementation. However, due to the need arising to use Tamil on computers, the local developers have started providing font based solutions, which resulted into creation of non-portable data. In order to encourage and enable data portability, the Govt of Tamilnadu have prescribed an ordered set of glyphs as standard based on the Tamil script. Further, the Govt of India have also prescribed glyph based standards for TTF fonts.

For the power hungry digital devices such as Pagers, Govt of India has prescribed a standard called Indian Standard Code for Language Pagers (ISCLAP). As the font technology has advanced to handle pre-defined character combinations, positioning of glyphs etc, Unicode standard got implemented on Operating Systems and application software. Eventhough, Unicode for Tamil is also based on the Tamil Script, it is getting implemented in Operating Systems and Application software as an International standard, which has enabled the shrink wrapped software developers to develop methods to handle world languages for their multilingual software products.

Standards prescribed for typing Tamil letters

There are Three standard typing methods prescribed for Tamil script 1. Typewriter, 2. Inscript and 3. Tamil '99. While Typewriter layout is based on letter formation using glyphs, Inscript and Tamil '99 are based on the phonetic letter formation.

Typewriter layout

Eventhough, the typing is largely based on the appearance of the letters, for some letters it is not so. For example, the zero width glyphs such as pulli, vowel signs ீ and ு are typed first before typing the base consonant. As per the Unicode standard, the vowel signs have to come (to be typed) after the base consonants. But, while typing Tamil text, some vowel signs like ீ and ு have to be typed first but have to be displayed to right side of the base consonant. For some other vowels like ஶ and ஷ, the vowel signs have to be typed first and then the base consonant to be typed. However, as per the Unicode standard, the vowel signs have to be placed after the consonant. For vowels like ஔ, ஓ and ஔள, the vowel sign have to be split and typed on left and right side of the base consonant.

Inscript and Tamil '99 layout

As these typing layouts are based on the phonetic concept, the vowel signs are typed after typing the base consonant. However, while editing, typing vowel signs contradicts with their visual appearance. For example, eventhough the vowel sign for vowel ஶ appears to the left of the base letter, the typing of vowel sign while editing has to take place towards right side of the base letter.

Difficulties in implementing tamil script

Tamil script has 247 Tamil letters for writing text in Tamil script. Obviously, it is not possible to accommodate all the Tamil letters on the computer keyboard for want of more keys. In order to implement Tamil script on computers, one has to think of possible means of combination of characters or glyphs to form all the letters. In some keyboard layout, such as Typewriter keyboard layout, it is not possible to provide feedback for all the keys to the typist while maintaining the use of aesthetic glyphs for forming vowelised consonants of இ, ஈ, உ, ஊ.

Such typing of vowel signs imposes difficulties while editing the text. In order to correct the spelling mistakes introduced due to placement of vowel signs by the Unicode standard, the developer has to make use of Private Use Characters (PUA). For such vowel signs, to avoid formation of composite letters is combination with the previous unintended base consonants. While editing the Tamil text, backspace deletes the vowel sign of the composite letter when the user intends to delete the consonant. For example, to delete க in கே, when the user presses backspace, instead of க gets deleted the vowel sign ே gets deleted. Such behaviors of the software annoy the user. As a result the editing process becomes cumbersome.

While implementing ISCII and Unicode, the basic requirement is to process combination of characters to form composite letters. While ISCII for Tamil has never been implemented in any of the Operating Systems, Unicode is being implemented in many Operating Systems such as Windows XP, Windows Vista etc. However, ISCII for Devanagari was implemented by IBM and APPLE Computer Inc.

Difficulties faced with Tamil '99 keyboard layout

The key feature of Tamil '99 keyboard layout is auto pulli. Eventhough, the auto pulli feature is an added advantage in reducing the number of keystrokes required to type the given Tamil text, it introduces errors in text while typing. It is learnt from the users that the users are comfortable in typing pulli than getting the same automatically. However, when the auto pulli has to be avoided in words like ழழழ, ந்த்தை inflected forms of காரர், the auto pulli feature introduces errors. When the user wants to avoid

auto pulli, he has to type அ immediately after the first letter. While typing being a subconscious skill, the users are forced to be conscious of words or letter combinations being typed.

Script reform

It is accepted by the archeologists involved in script deciphering that the evolution of script started first from logographic writing system then improved over to syllabic writing system and finally to phonemic writing system. Tamil writing is neither syllabic nor phonemic but it has graphemes for phonemes and orthographic syllables. The secondary forms of vowels are written to right, left and both sides of the base consonant. Vowel signs of இ, ஈ, உ, ஊ combine with base consonants to form vowelised consonants, which are designed to be a single glyph. Such orthographic formation leads to difficulties in implementing editing text.

Writing system largely depends on the material used for writing and the writing instrument. When Tamil script implementation is considered for computer, the material used for writing becomes the storage device where Tamil letters are stored in its encoded form. The writing instrument to be considered for computer is keyboard. Even otherwise, if the writing system using pen is considered, it would be convenient to have a representative glyph to represent the Vowels இ, ஈ, உ, ஊ.

The history of writing establishes that revision of writing taken place in 1. Adopting different script than the existing script as in Malay, 2. Corrections introduced in the existing letters as in Chinese and Malayalam, 3. Increasing or Decreasing the number of letters as in Kannada.

Reform in Tamil script writing is required to improve the efficiency of composing Tamil letters and editing Tamil text. As the script reform proposes to introduce graphic shapes as graphemes for vowel signs of vowel உ and ஊ, it enables to provide feedback while typing Tamil text and allows to edit the text like deleting visible vowel signs instead of deleting invisible vowel signs.

Required vowel signs for script reform

In Tamil script, additional letters or signs are formed, sometimes by adding a closing circle at the end of the stroke as in ீ, ெ signs and letter ூ or by adding additional stroke to mark elongation as in letter ே. It is also to be noted that the signs added to mark vowel signs of உ, ஊ for the Grantha letters is contradicting and thus leads to confuse the users. As a result, the ு sign is normally mistaken for ூ sign because of the rounded circle used at the end of the stroke in ு sign.

In order to achieve script reform in Tamil, vowel signs are required to represent the vowels இ, ஈ, உ and ஊ. For இ and ஈ the existing signs ி ீ could be modified to use it in standalone form. Vowel signs for உ and ஊ have to be based on the familiar sign already used to form vowelised consonants.

Suggested vowel signs

In order to achieve the script reform in Tamil, the existing signs for இ and ஈ could be used with little modification as ி and ீ so that these signs could be written to the right side of the base consonant. On the basis of hand movements used for writing vowel signs and the kind of stroke already familiar to Tamil users, the following vowel signs are suggested for vowel signs of உ and ஊ respectively:

ு ூ

The above vowel signs are written with the same clockwise hand movement followed for writing other vowel signs like ி, ீ, ூ and ௃. Also, such type of vowel signs are used for some vowelised consonants formed with the combination of ் and ுஹ vowels.

Benefits of script reform

The script reform adds the following benefits to Tamil script usage:

1. Reduces the number of glyphs by 72, which are required to form the vowel consonants with ி, ீ, ூ ௃ vowel signs.
2. Makes typing and editing Tamil text easier.
3. Enables feedback for typists.
4. Since the number of glyphs required for Tamil script reduces to 1/3, it would be easier to implement Tamil script in every digital gadgets wherever English is implemented.
5. It would enable development of application software and games as available in English.

Suggestions

- Vowel signs have to be allowed to be typed in standalone form inspite of its dependent vowel status. Vowel Signs have to be allowed to appear without its preceding dotted circle when typed as a standalone characters.
- The overhead of placing the vowel signs to the right of the base consonant has to be moved to the application software through normalization and convenience of typing the vowel signs as per their appearance has to be enabled in the application software. This is suggested as a feature for Tamil enabled software and this feature has to be considered while testing the software prior to certification of the software.
- Backspace key has to delete the immediate visible letter, instead of deleting the vowel signs of the composite letter.
- Auto pulli feature has to be removed from the Tamil '99 keyboard layout to enable error free typing.
- Two new symbols are suggested to represent vowel signs of ் and ுஹ to enable script reform.

Conclusion

An attempt has been made to study the present status of Tamil script in digital media and the associated issues. As a remedial action script reform has been discussed and some suggestion are made for consideration. It is believed that the factors discussed and he suggestions made will help to improve the implementation of Tamil script in digital media.



இணையம் வளர்க்கும் தமிழ்

www.tamilinternetconference.org