

**t.n.c.venkata rangan**

**CMD, Vishwak Solutions Pvt. Ltd.**

Former Chair, International Forum for Information Technology in Tamil



# Relevance of Unicode to e-governance

*India has 122 major languages and 2371 dialects*



# Linguistic Diversity in India

- According to Census 2001 India has 122 major languages and 2371 dialects
- One Language –many script
- Many Language –one script
- Out of 122 languages 22 are constitutionally recognized languages

# Challenges

- Lack of availability in local language has been a reason for slow progress of ICT & benefits not percolating down to common man

The screenshot shows a government website interface. On the left, there are links for various registration services: Document Registration (Deed/PowerofAttorney/Lease etc.), Marriage Registration, Nursing Home Registration, Partnership Firm Registration, and Sales Tax Registration. Each link includes an FAQ and a download of forms/affidavit. On the right, there is a section titled 'Online Submission of Applications' with a form for applicant details (Name, Sex, House No., Street/Sector/P.O. etc., Village/Locality\*) and a list of certificate types to apply for online, such as Registration of Marriage Certificate, Schedule Caste (SC) Certificate, and others. A 'Click Here To Apply Online' button is at the bottom.

The screenshot shows a 'PERSONAL PARTICULARS FORM (In Duplicate)' for passport application. The form includes fields for:
 

- 1. Full name (Initials not allowed)
- 2. Sex: Male / Female / Others
- (a) Has the applicant ever changed name?
  - (b) If yes, previous name:
- 4. Date of Birth
- 5. Place of Birth
- 6. Profession
- 7. a) Father (Surname) (Name)
- b) Mother

 A note on the right side of the form states: 'Paste your cross signed recent colour photograph, size 3.5x5.5 cm'. The website header is in Hindi, and there are navigation menus on the left and right.

# Encoding - Introduction

- Computers at their most basic level just deal with **numbers**
- Store letters, numerals and other characters by assigning a number for each one.
- Single 8-bit characters sets, limited us to 256 characters max.

## Why a Standard Encoding?

- Ensure sharing of information
- Help seamless interoperability of data across e-Governance applications
- Developers from any part of the globe could interact through the application

# What is Unicode?

- Unicode was initiated in **1987** by Joe Becker from Xerox and Lee Collins and Mark Davis from Apple
- Provides a unique number for every character, for any
  - **Platform**
  - **Program**
  - **Language**
- Visual rendering (size, shape, font, or style) is done by other software, such as a web browser or word processor
- 107,000 characters covering **90 scripts**
- Supported by other standards bodies including ISO, W3C, IETF, ELRA and BIS

# Adoption by GoI in E-Governance

- All 22 Languages including Tamil has represented and included in UNICODE by TDIL, Govt. of India
- In Nov 2009, Expert Committee on Technology Standards for Indian Languages & National e-Governance Plan (NeGP) under DIT has notified in **Gazette of India** that “**Unicode 5.1.0 and its future versions as standards for e-Governance Applications**” - <http://bit.ly/u-expert> & <http://bit.ly/u-goi>

# Adoption by states in E-Governance

- In August 2008, Kerala Government issued a GO on use of Unicode in Government Offices.  
<http://bit.ly/u-gok>
- Recently, West Bengal Govt. and SNLTR has decided to accept UNICODE 5.0 and upwards as the standard for data representation and encoding for Bangla. <http://www.nltr.org/SNLTR/>
- In June 2010, **Tamil Nadu Government** has announced “Unicode shall be the main 16-bit encoding to be used in all applications where support for Tamil is available” <http://bit.ly/u-gotn>
- On 27 Jul 2010, Maharashtra state Chief Minister Ashok Chavan announced government will shortly introduce Unicode system for easy access of Marathi language across the world



# Road Ahead

- Take stock of all existing legacy, proprietary, 8-bit encodings used in Data, Documents & Applications
- Plan and execute to convert to Unicode
- Develop & Procure all new systems & applications in Unicode
- Ensure all new documents and data are in Unicode
- Ensure Unicode & Indic Languages are supported in all PCs, Mobile Phones & Embedded devices

ధన్యవాదము

ధన్యవాదము

అలా

ధన్యవాదము

నిండేటట్లుగా

నూతన

ధన్యవాదములు

అభినందనలు