Tamil Language Computing and IDN

S. Maniam

Board Member

INFITT

A MINC Affiliate

உத்தமம்
INFITT
உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம்
International Forum for Information Technology in Tamil

MINC

Background

- 1.5 + Billion population
- Over 600+ languages (many more dialects)
- Many with their own scripts.
- 18 official languages on rupee currency note
  Hindi, Tamil, Urdu, Orissa, Gujerati, Marati,
  Kannada, Telugu, English etc.
- Contrary to popular belief only 125 M speak English.
- Largest is Hindi with   300M (+ 10M)
              Urdu with   130M (+160M)
              Tamil with   70M (+   6M)
              Telugu with   55M (+   3M)

Issues – Diversity & Overlaps

- **Urdu**
  3rd largest non-latin script
  Pakistan (150M),India (130M)
  Arabic and Indic Script (but different unicode pts)

- **Tamil diaspora is world-wide**
  Mainly Tamil Nadu State in India (70M)
  National Language – Singapore, Sri Lanka
  Official Language – India, Mauritius, Malaysia
  Significant Use – UK, USA, Canada, Australia,
  Germany, Switzerland, HK,
  South Africa.
- **30+M Indians outside India**

Problems

- **Each of main 18 languages had about 32 different encodings of the script.**
- **India has about 20 main states, with many of the main languages corresponding one to each state.**
- **The Federal Government's Advanced Computing Institute C-DAC developed one standard encoding for each language/state (1999).**
- **C-DAC was ignored by the states and their chief ministers.**

Progress

- **INFITT and a pre-INFITT group, spearheaded and now leads the effort to standardize the encodings.**

- **INFITT is the International Forum for Information Technology in Tamil (infitt.org)**

- **Formed in 2000 based on a pre-INFITT group, partly with input from MINC.**

- **Worldwide group with 13 countries – India, Singapore, Malaysia, UK, USA, Sri Lanka, Germany, Switzerland, Mauritius, South Africa, Canada, HK, Australia.**
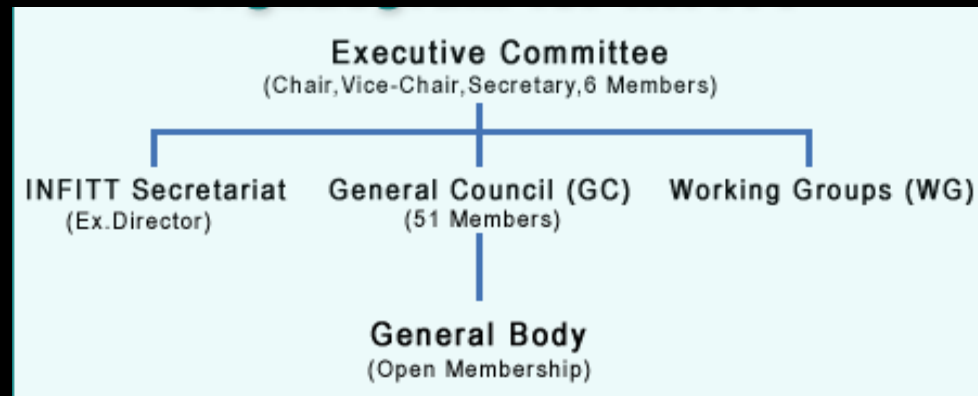

உத்தமம்
INFITT
உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம்
International Forum for Information Technology in Tamil

MINC

Progress led by INFITT

- INFITT helped create the TAB/TAM standard taking the best from the 30+ other encodings and the 6 governments that have Tamil as a national/official language endorsed it via local steering committees in 2000.
- INFIIT pushed Tamil as the first Indian language to be put in Windows 2000.
- Popularity and use is steadily growing while prior encodings are dying.
- Based on the INFITT model, two other Indian languages and their state governments have been officially standardized in encoding – Hindi and Telugu.
- Both are in Windows XP.

INFITT Structure

Executive Committee
(Chair, Vice-Chair, Secretary, 6 Members)

INFITT Secretariat (Ex.Director)

General Council (GC) (51 Members)

Working Groups (WG)

General Body (Open Membership)

- **INFITT has 51 council members and 7 working groups ranging from Unicode to Linux.**

INFITT Outreach

- INFITT has annual conferences with reviewed papers, presentations, publications, posters and workshops since 2000

  Singapore   2000 (85 000 attendees, 500 delegates)
  Malaysia   2001 (100 000 attendees, 800 delegates)
  San Francisco   2002 (500 attendees, 150 delegates)
  Chennai, India   2003 (10 000 attendees, 500 delegates)

- INFITT events have been launched by

  Prime Minister   M Mahatir   of Malaysia
  President   SR Nathan   of Singapore
  Chief Minister   Karunanithi   of Tamil Nadu, India

INFITT IDN Efforts

- IDN working group was initiated in 2000. I am Chairman (S. Maniam).
- Tab/Tam was selected as the encoding for Tamil and endorsed by all countries.
- Windows support was arranged in 2000.
- The IDN gTLDs equivalent for .com, .net, org, .edu & .gov were finalised as shown below

Com	வணி
Net	இணை
Org	அமை
Edu	கல்வி
Gov	அரசு

IDN Deployments

- These gTLDS were commercially deployed in Tamil Nadu State in India in 2000 with state support.
- The growth has been slow, owing to total internet users in India is only 12 million &Tamils < 1M. Many of these early-users already fluent in English.
- But the need is increasing.
- Hindi and Telugu were commercially deployed in 2001, by Information Minister and Deputy Prime Minister of India and by the Chief Minister of Andhra Pradesh, respectively.
- Like Tamil, growth slow but increasing need.

INFITT IDN Effort and MINC

- **INFITT has been involved with MINC since 2000.**
- **In 2002, the INFITT IDN working group announced a cooperation agreement between MINC and INFITT, to participate in the inter-operability testing for IDN deployments pursuant to the final IETF IDNA standard.**
- **INFITT submitted an RFC to MINC for the Unicode Language Code Point tables in April 2002. Tamil is distinctive in that its code-points do not overlap or conflict with any other script.**
- **Next, we are planning to study/develop a dispute policy and prohibited word list in conjunction with MINC.**
- **INFITT intends to participate aggressively and as a leader in the IDN movement to keep a 4000+ year old language not only alive, and current but also leading the way.**

Summary

The mission of INFITT is to facilitate the ancient Tamil language and its sister Indian languages to remain current, vibrant and modern by keeping up with global IT developments, particularly in English.

http://www.infitt.org

http://www.minc.org

S. Maniam <s_maniam@pacific.net.sg>