

Liaison Report INFITT WG02 (Unicode Tamil Segment)

1) TI2002 (Tamil Internet 2002)

The International Forum for Information Technology in Tamil (INFITT) officially announced that the fifth Tamil Internet international conference and exhibition (TI2002) will be held at Foster City, San Francisco Bay Area, California from September 27 - 29 this year.

This event is organized by INFITT in close collaboration with the Center for South Asia Studies and the Chair in Tamil Studies of the University of California, Berkeley. Conference web site is:

<http://www.infitt.org/ti2002/>

The prior year's conference was held in Kuala Lumpur, Malaysia. I spoke there and I can attest an intense amount of interest in Unicode (they actually asked me to hold a special extra Q&A session about Unicode and Tamil which picked up most of the attendees of the conference. There is every indication that the people who come to a conference in Silicon Valley will be even more interested (and of course more knowledgeable/more likely to be running systems that can use Unicode now).

2) M.Ponnaivaikko, is the director of the Tamil Virtual University in Tamil Nadu (website at <http://www.tamilvu.org/>). It is an organization that provides a lot of information on Tamil language and culture. They are actually joint Associate members of Unicode with the Government of Tamil Nadu.

M. Ponnaivaikko has been behind a much touted study that was designed to compare Unicode Tamil with other encoding schemes, in part to allow Tamil Nadu to answer the requests of the Government of India for member states to provide information on the suitability of Unicode. The purposes of this study are best summed up as follows:

There has been a uniform feeling in the minds of the Tamil computing persons in the world that the present Unicode standard for Tamil is not adequate for efficient and effective usage of Tamil in computers, particularly, for applications like Database Design with search and sort facilities and Natural Language Processing. While considering an alternative to the present Unicode Standard for Tamil, there are two different views : One, the scheme that incorporates only the vowels and consonants into the encoding scheme, leaving the other characters to a rendering process. Two, the scheme that incorporates all the Tamil Characters including vowels, consonants and vowel-consonants as characters in the standard.

The study was funded by the government of Tamil Nadu. The areas tested in the study:

- (i) Space occupation of the Encoding Schemes.
- (ii) Efficiency of Text Processing.
- (iii) Efficiency of Database Application
- (iv) Efficiency of Morphological Analysis.

The summary document describing results is provided here also (and is publicly available as the INFITT WG02 list is open for all to read, though only members can post to it). The conclusion (as you might have guessed):

...the discussion presented above the Scheme 3 proves to be efficient in terms of the resource consumption and execution time. Hence, Scheme 3 is recommended for incorporation in Unicode Standard for Tamil.

Source data and information about methodology has been promised, although has not yet arrived (the study as completed in the end of May 2002). Several people on list pointed out flaws from what they could review, and there has been no discussion about the study at all since May. M. Ponnaivaikko will be speaking at TI2002 in Foster City (see above) about the study and its conclusions and people in the area may wish to seriously consider attending to see this presentation.

3) In the interests of time, INFITT is going to present to the Book committee a detailed list of issues to add, remove, and edit for the Tamil block description. This has been promised by the end of the month.

4) After recent discussion on the Unicode List started by Mr. Sinnathurai Sivas for "Unicode Tamil", a plan which has evolved somewhat over the last few years but is still not being widely looked at seriously by people in the community, the Executive Director of INFITT (Arun Mahizhnan) communicated an official position on the matter, which Sivas agreed to abide by for the sake of the community. That post is here quoted in full:

Dear Members and friends of INFITT,

Allow me, as the Executive Director of INFITT, say a few words on the possible contributions INFITT could make in this regard.

As Mr Manivannan has pointed out, INFITT is not a standard-setting body but an advisor to governments and standard-setting bodies, and a consensus builder amongst our community. It also has a role as path-finder and in that sense it is possible for INFITT to explore new ideas and processes to improve the language from a technological standpoint. However, INFITT's credibility depends much on its ability to marshal the necessary resources to make our case and equally importantly, on our ability to represent a consensus. Those two factors oblige us to adopt a certain modus operandi: first we need to set up an experts group to study the problem and come with possible solutions. To that end, in the past, we had set up Working Groups or Task Forces. The Executive Committee of INFITT prefers this as an effective way to problem-solving. And the EC is the only group within INFITT which can set up such WGs or TFs. Therefore, if any or some of you wish INFITT to pursue this idea that Mr Sinnathurai Sivas is proposing, do send a formal request to the Secretariat (i.e. me or Nara). We are happy to process such requests.

The other factor, mentioned earlier, is critical too. If members of INFITT wish the organisation to represent them, then they should desist from taking personal petitions to the authorities even before INFITT has had a chance to deliberate on the matter. For example, we have a Working Group on Unicode, trying very hard to sort out some issues with Unicode. If a member of INFITT chooses to go directly to Unicode with her/his own recommendation, even as we are grappling with a solution, it will undermine the image of our organisation. Of course, we could and would take counter measures, but it just soaks up so much of our precious and voluntary time and effort. So it is imperative that INFITT members observe this basic obligation to the organisation to play by the rules.

The above does not, however, prevent anyone from expressing her/his personal views, however contrarian, within our lists. In fact, the more we rigorously and robustly debate the issues the better. The above also does not prohibit anyone from expressing her/his personal views in other lists or to anyone else. No one can stop you from doing that. However, do remember that the more conflicting views going out of our community, the more difficult to convince outsiders that INFITT represents a consensus. So we urge you to exercise your good judgment in deciding how far you want to go it alone.

If the above conditions could be met, INFITT, I'm sure, will be happy to consider any serious request from its members to serve our language better.

Warm regards.
Arun Mahizhnan