Report of the South Asia subcommittee
UTC meeting # 209
August 6-10, 2007

This document is the report of the meeting of the South Asia subcommittee, held on August 6. It represents the consensus of the participants, formulated as recommendations to the UTC; those positions should not be construed as UTC positions.

1. Chillus accepted in Amendment 4

Recommendation:
- continue to support the encoding of those 6 characters
- encourage the continuation of Cibu's work clarifying the representation in Malayalam, both before and after the acceptance of those 6 characters
- keep the name of *U+0D7C as MALAYALAM LETTER CHILLU RR, with an annotation to the effect that this is also writes the chillu form of RA

2. Other additions in Amendment 4, for Oriya and Malayalam (UTC agenda item C.20.1, L2/07-196)

Recommendation:
- continue supporting those additions, without change

3. Meitei Mayek (UTC agenda item C.2, L2/07-256)

Recommendation:
- incorporate the comments of L2/07-256 in a revision of the proposal (this revision will help the writing of the block description)
- reaffirm the position taken by WG2 to not include danda characters

4. Malayalam YA chillu (UTC agenda item C.20.2, L2/06-337, L2/07-242, AI 109-A75)

Recommendation:
- both proposals rely on the same single source. The subcommittee would feel more comfortable with evidence from multiple sources. Also, the subcommittee would like to understand better the range of glyphs for the chillu YA

5. Malayalam TA chillu (L2/06-343, AI 109-A75)

Recommendation:
- since the shape of the chillu TA is the same as that of the chillu LA, it seems that there is a single abstract character with different pronunciations; this is best accounted for by annotating U+0D7D MALAYALAM LETTER CHILLU L to that effect.

6. Kashmiri (L2/06-372, AI 109-A7)

Recommendation:
- nothing to do beyond the completion of the open action item. When the information is finalized, action item to Eric to communicate that to the CLDR committee for appropriate update of the examplar characters in CLDR

7. North Indic Number Forms (UTC agenda item C.17, L2/07-139, L2/07-238)

Recommendation:
- to accept the changes proposed in L2-07/238 as is (changes to character names, creation of a new block)
- to recommend to WG2 the inclusion of those characters in Amendment 5, if possible
- to ask Anshuman Pandey to produce another revision of L2/07-139R including those changes

8. Bengali Ganda currency mark (UTC agenda item C.1, L2/07-192)

Recommendation:
- to accept this proposal as is and to recommend to WG2 the inclusion of that character in Amendment 5, if possible

9. Additional Devanagari character (L2/06-373, AI 109-73)

When the UTC looked at L2/06-373, it was asked whether the U+02BC MODIFIER LETTER APOSTROPHE was appropriate for the uses described in the document. Swaran Lata reported that this is indeed appropriate.

Recommendation:
- add annotation to that effect to U+02BC
- give an action item to the Govt. of India to provide a draft update to the Devanagari block description

10 . TACE-16

The discussion resulted in a draft "charter" for what the subcommittee should do regarding TACE-16.

Goal: ensure that Unicode meets the needs for representation and processing of Tamil.

This may or may not require the encoding of new characters. Any recommendation should exhaustively examine the implications, including on existing data, on existing software (processing, display, etc), on education about the standard, on consistency of model for the Indic and other South Asian scripts.

The scope of the subcommittee is to review the issues and to make recommendations to the UTC.

Step 1: Identification of the issues

Identify the issues (problems or perceived problems) with the current representation. Determine whether they are issues with the standard itself (encoding, properties, or algorithms) or with implementations. Determine the nature of the issues: technical, perceptual or educational.

Candidate issues:
1 disconnect of the code chart with the user expectations
2 efficiency in storage/transmission
3 efficiency in processing
4 correctness of implementations
5 difficulty of implementation

Step 2: Evaluation of possible approaches

This enumeration of possible approaches does not preclude the examination of other approaches (which may extend on or combine the approaches below). The questions listed for each approach are illustrative of the kinds of questions that need to be answered for a proper evaluation of the approach; they are not exhaustive.

Approach A: current model

How would those issues be addressed with the current representation? Are there any enhancements (new characters, changes to properties, addition of properties, guidelines, documentation in the standard) that would alleviate those issues?

Approach B: TACE-16 repertoire added to Unicode

How would adding the TACE-16 repertoire to Unicode address those issues? And what would be the new problems created by the introduction of that repertoire?
For example:
• dual encoding and stability policy
• does it need to be in the BMP, and if so, how does it fit there?
• would encoding in a non-contiguous area help or hurt compression techniques?

Approach C: TACE-16 repertoire in the PUA

What are the issues that applications are faced with?
For example:
• collisions with other well-established PUA uses, such as CJK:
  - there is not always an "official" mapping, different vendors do different things
  - PUA conflicts:
      HKSCS 9571 (U+2721B) → U+E78D
      GB18030 A6D9 (,) → U+E78D
  - PUA differentiation:
      HKSCS 8BFA (U+20087) → U+F572
      GB18030 FE51 (U+20087) → U+E816
• PUA characters cannot be used in IDN.

Approach D: TACE-16 as a separate IANA-registered character set

How simple is it to add support for a new character set (with a well-defined mapping to the existing Tamil block) to exisiting Unicode-based applications? Can this be done in a timely manner, across enough products to achieve viable workflows? What are the implications for already shipped software?

Recommendation:
- the UTC adopts the text above as the charter to guide future discussion by the South Asia sub-commmittee on Tamil issues.

11. Vedic (UTC agenda item C.12, L2/07-230, L2/07-254)

The South Asia subcommittee is very pleased with the progress on the Vedic encoding proposal.

Recommendation:
- the UTC would like to see a revised proposal, which addresses the comments received during discussion and incorporates contributions from the Government of India. This revised proposal should be on the agenda for a future UTC meeting. While the progress is very significant, we believe this is not mature for inclusion in a ballot. We expect the revised proposal to be ready for balloting.

12. Kaithi (UTC agenda item C.5, L2/07-199)

Recommendation:
- give an action item to Anshuman Pandey to revise the document to account for the comments made in discussion (Deborah Anderson will communicate those comments).

13. ZWJ and ZWNJ in identifiers (UTC agenda item B.11.3, L2/07-255)

The subcommittee discussed the issue and felt that more examples were required, as well as more clarification on the question, and deferred the topic to the UTC plenary.

Participants:

Deborah Anderson
E. K. Bharat Bhushan
Peter Constable
Mark Davis
John Jenkins
Michael Kaplan
Swaran Lata
Rick McGowan
Eric Muller
Lynn Ruggles
Murray Sargent
Peter Scharf
Michel Suignard
V.S.Umamaheswaran
Ken Whistler