

- Towards development of Script Core Technology a Glyph Property Editor and a Rule Engine has been completed; and
- Work is in progress for Urdu Spell Checker, Urdu-Hindi and Hindi-Urdu Dictionary, Muslim names and Technical Dictionary, and Converters.

14. RCs - ILTS : JNU : Chinese, Japanese & Sanskrit

- Teaching learning system for Sanskrit under development. 5 lessons with multi-media support designed & implemented;
- Data structure for Sanskrit lexicon designed. Implementation and data entry being progressed;
- A web site to facilitate Chinese tourist being progressed; and
- Web lessons for learning Chinese and Japanese languages are being progressed.

Impediments in IT Localization & Penetration

- Lack of industry involvement due to constrained demand;
- Subcritical and Unsustained demand in BIMARU & other States;
- R&D in language technology so far open ended, not product-driven in time-bound manner;
- Negligible software tools and re-usable components in public domain;
- No incentive for Computer Scientists working in the area of Indian Language Processing and also limited scope of their career advancement;
- No formal IT courses in the curricula for linguistics, language teaching, and language studies. Lack of IT culture in language graduates;
- Unable to check import of IT products and services which don't support Indian language(s);
- No consensus on National Standards: ISCII-88, ISCII-91, UNICODE, and other propriety codes;
- Content is largely glyph-coded not (ISCII) character-coded;
- Slow pace of transfer of language technology. No strategy for language technology marketing;
- Industry consortium for language technology yet to form.

Feedback on UNICODE Standard 3.0

- UNICODE is emerging as international de-facto industry standard. It is 2-byte code. The Unicode consortium provides a number of on-line resources for obtaining information and data about the Unicode standard, as well as updates and corrigenda. These are:
Unicode Web site: <http://www.unicode.org>;
Unicode Public Mailing List: unicode@unicode.org
- The following are of our relevance in the UNICODE standard version 3.0.
 - (a) IPA extensions (0250-02AF) on PP 359-367
 - (b) Devanagari (0900 - 097F), Bengali (0980 - 09FF), Gurmukhi (0A00 - 0A7F), Gujarati (0A80 - 0AFF), Oriya (0B00 - 0B7F), Tamil (0B80 - 0BFF), Telugu (0C00 - 0C7F), Kannada (0C80 - 0CFF), Malayalam (0D00 - 0D7F), on pp 211-235 and 401-425
 - Observations and comments :
 - (i) Sinhala (0D80 - 0DFF), Tibetan (0F00 - 0FBF), and Myanmar (1000 - 109F) follow the alphabetic order similar as in DEVANAGARI and other Indian scripts.
 - (ii) Elaborate description for DEVANAGARI (U+0900 - U+097F) is given on pp 211 - 223. This needs review in view of some additions/revisions. This includes sections on Standards, Encoding Principles, Rendering mechanism, Consonant letters, Independent Vowel letters, Dependent Vowel signs (matras), Viram, consonants conjuncts, Rules for Rendering (R1 - R14), Memory Representation and Rendering order, sample half-forms and ligatures, combining marks, digits and punctuation symbols.
 - (iii) Use of VIRAM in Unicode 3.0 is wrong. This must be replaced by HALANT. VIRAM means full stop as punctuation mark. Halant suppresses inherent vowel, and thus used with consonant to represent pure consonant.
 - (iv) Rendering mechanism is suggested as follows :

(a) Linear conjunction

क	्	व	=	क्व
0915	094D	0935		
क	्	ष	=	क्ष
0915	094D	0937		
त	्	र	=	त्र
0924	094D	0930		
ज	्	ञ	=	ज्ञ
091C	094D	091E		

(b) Vertical conjunction

क	ZWJ	व	=	क्व
0915	200D	0935		