

- **Tools and Environments:** There are very few shareable tools and environments for Language Engineering. One example is the GATE (General Architecture for Text Engineering) effort at Sheffield. There is scope to work together in adding South Asian language support into such tools and environments, and also to broaden the scope of such tools, by adding components specific to these languages, and also application areas such as translation and generation. Evaluation and comparative study of existing tools and environment should be made.
- **Applications:** The major application areas of Language Technology are Information Retrieval and Extraction, Translation aids, OCRs, spellcheckers and speech processing systems. Some of the lesser known, but very relevant applications include localization, language learning, and communication tools for the blind, the illiterate and the semi-literate. The specific modes of cooperation in these areas needs to be explored further.
- **Standardisation:** There is an urgent need for India to actively promote and participate in standards building activities. There are four aspects or levels of standardization — Encoding (characters, fonts etc), Representation of text (logical organization), Linguistic Annotation, and Access (indexing and metadata). India should participate in global initiatives such as the TEI (Text Encoding Initiative) and OLAC (Open Language Archive Community), and make sure that its unique language standards needs are addressed, since it is not necessary that methods used for other languages are directly applicable to South Asian languages. For example, there are some unique features of South Asian languages, such as their phonetic nature and presence of composite glyphs, which makes it necessary to handle them specially at the encoding level. Similarly, due to their rich morphology and relatively free word order, they may need different kinds of analysis, representation and tagging models.
- **Evaluation:** Systematic and objective evaluation of NL technologies and products, though not easy, is a necessary and effective mechanism to establish and extend the state of the art. Evaluation can take three forms, depending on how mature and established a technology is
 - a. Technology Oriented Evaluation, which typically involves voluntary participation of groups with focus on specific components or techniques, not necessarily complete products.
 - b. Product Oriented Evaluation, which involves user feedback on actual systems.

- c. Certification by some authority as to interoperability, usability etc. It was felt that Indian NL groups should get together and try to organize Technology Evaluations, along the lines of the TREC and MUC conferences.

New Software Testing Facility

Under Soft QC/QM project of Computer Development Division of Ministry of Information Technology Electronics Test & Development Centre, Thiruvanniyur (Near Lattice Bridge), chennai-600041 has established a software Testing & Training Laboratory (STTL) at chennai. This laboratory is capable of testing software products as per ISO/IEC 12119 standard. Academia, Industry and Language Technology projects being carried out under MIT shall benefit from this facility.

Contact Person :

Mr. R. Subburaj, Director ETDC, Chennai
E-mail : etdccn@md3.vsnl.net.in Tel : 044-4928192

Now Domain Names in Regional Languages

- US based domain name Registration Company, Network Solutions (<http://www.networksolutions.com>) announced domain name registration service in nine Indian regional languages. They are launching domain name registration services in nine different Indian Languages-Bengali, Gujrati, Hindi, Kannada, Malyalam, Oriya, Punjabi, Tamil and Telugu to enable people to break the language barrier. Besides bridging the digital divide the expansion of multilingual services would induce small and medium businessmen to conduct transactions online in their language. Until recently every domain name registration was done using exclusively ASCII characters. India ranks third in the domain name registration in the Asia-Pacific market after Korea and China. Along with these languages one can use .org, .net and .com for Internet Domain Name Registration. The invitation price for the service initially would be \$25.
- VeriSign Global Registry services (<http://www.verisign.com>) which maintains the master database for the majority of the World's domain names has decided to accept names in the simplified and traditional Chinese, Japanese, and Korean characters. Next, VeriSign plans to accept domain names in Arabic, and names with Spanish and Portugese characters not used in the English.