



### 3. Calendar of Events-Year 2002

- LRMT-2002 : NLP workshop on developing language resources for machine translation at IIT Hyderabad, 17 June-6 July, 2002 (with optional WSD Workshop : 8 July to 13 July, 2002 .  
Web : <http://www.iit.net>
  - Indo-European Conference on Multilingual Communication Technologies at C-DAC Pune, June 24-25, 2002.  
Web : <http://www.cdacindia.com.iemct>
  - The Second International Natural Language Generation Conference (INLG2002), Arden Conference Center, Ramapo mountains, New York City, USA, July 1 to 3, 2002.  
Web: <http://www.research.att.com/~rambow/inlg.html>
  - Student Research Workshop, July 6-11, 2002, Workshop on Word Sense Disambiguation : Recent Successes and Future Directions, July 11, 2002, Workshop on Effective tools and methodologies for teaching NLP and CL, July 7, 2002, Workshop on Speech-to-Speech translation : Algorithms & Systems July 11, 2002, Workshop on unsupervised Lexical Acquisition, July 12, 2002, at ACL-02, Philadelphia, Pennsylvania, USA.  
Web: <http://www.acl02.org>
  - 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7 – 12 July, 2002.  
Web: <http://www.acl02.org>
  - Summer Workshop on Language Engineering, to be held in Baltimore, MD, USA, from July 8 to August 16, 2002.  
Email: [sec@clsp.jhu.edu](mailto:sec@clsp.jhu.edu)
  - Natural Language Processing in the Biomedical Domain, University of Pennsylvania, Philadelphia, PA from 11-12 July, 2002.
  - Lexicom@ITRI Workshop will be held at Information Technology Research Institute, University of Brighton, England from 14-19 July 2002.  
Web : <http://www.itri.bton.ac.uk/lexicom/>
  - 19<sup>th</sup> International Conference on Computational Linguistics, Howard International House, Taipei, Taiwan, August 24 – September 1, 2002.  
Web: <http://www.coling2002.sinica.edu.tw/>
  - COILING 2002 3rd workshop on Asian Language Resources and International Standardization at Center of Academica Activities, Academia Sinica, Taipei, Taiwan, August 31, 2002.  
Web: <http://www.coling2002.sinica.edu.tw/>
  - Marathi Word Processing for Govt. Officials, IIT Mumbai, June 2002. Web : <http://www.cse.iitb.ernet.in/~pb>
  - Association for Machine Translation in America-2002, Tiburon, California , October 8-12, 2002.  
Web: [www.amtaweb.org](http://www.amtaweb.org)
  - UNL symposium along with a Machine Translation Workshop, IIT Mumbai, June 2002  
Web : <http://www.cse.iitb.ernet.in/~pb>
  - ICON (Indian Conference on NLP)-2002, NCST Mumbai, Dec. 2002. Web : <http://www.ncst.org>
  - A Workshop on NLP and Applications, Anna University, Chennai, Dec. 10-21, 2002. Web : <http://annauni.edu>
- Note :** ■ Indicates conferences in India.

### 4. Standardization

#### Revision of Unicode Standard-3.0 for Indian Scripts

Unicode Standards are widely being used by the Industry for the development of Multilingual Softwares. Indian scripts are also included in the Unicode Standards. Unicode consortium had taken the basic inputs for standardisation of Indian scripts from ISCII-1988 document. There are some deficiencies in the present Unicode Standards for Indian Scripts and need to be removed for proper representation of the Indian Scripts.

Department of IT is the voting member of the Unicode Consortium. The ministry has collected inputs for each of the Indian scripts in order to make a single unified presentation of Indian scripts to the Unicode Consortium. A number of meetings starting from November 2000 have been organized with the concerned State Governments/ Organizations/ Experts and Industry on this subject. Based on the discussions / feedback draft Code charts for each of the script have been prepared along-with code details. The first draft proposal of the proposed changes in the existing Unicode Standards for Indian scripts was published in the May 2001 issue of the *TDIL Newsletter- VishwaBharat@tdil* to get the feedback from experts/ industry / users working in the area of Indian Language Software Development. The news letter was also sent to the members of the Unicode Consortium for their comments and initial response. The draft proposal was discussed in the Unicode Technical Committee (UTC) meeting held in USA in November 2001. The minutes of the UTC Meeting are available in document No. L2/01-430R and L2/01-431R of Unicode Consortium. The URL for viewing these documents are as given below :

<http://tdil.mit.gov.in/newsletter1.htm>

<http://www.unicode.org/L2/L2001/01304-feedback.pdf>

After getting the feedback from state government / linguists / experts / industry the second draft proposal is also being prepared. The final draft of the Devanagari script was published in January 2002 issue of *tdil* newsletter *Vishwabharat@tdil*. We are thankful to user community for their valuable feedback. Please read character 'ज' as 'घ' at page number 27, 2nd column, line number 20.

In this issue of the TDIL Newsletter we are publishing draft for proposed changes in present Unicode Standards for Gujarati, Gurmukhi, Malayalam, Oriya & Telugu script. The whole write-up is divided into three parts:

1. **Code chart** - For quick reference of the characters and their code value.
2. **Code Details** - For reference of the character names and annotations.
3. **Script Details** - A review of the script to cover various technical characteristics of the script.