



1. Calendar of Events-Year 2002-03

- Association for Machine Translation in America-2002 Conference, Tiburon, California, Oct. 8-12, 2002.
Web : www.amtaweb.org
- National Seminar on 21st Century Reality : Language, Culture & Technology, MGIHU, Delhi, 29-31 Oct. 2002.
Web : www.hindivishwa.nic.in
- LRC-2002 : 7th Annual Localization Conference-econtent Localization, Dublin, Ireland, Nov. 12-13, 2002.
Web : www.localisation.ie
- International Conference on Universal Knowledge & Language, Taj Goa, Fort Aguada, 25-29 November, 2002.
Web : <http://www.cfilt.iitb.ac.in>
- 5th East Asia Forum on Terminology (EAFTerm 2002), HaiKou city, HaiNan Province, China, 8-11 Dec. 2002.
- Language Engineering Conference (LEC-2002), at University of Hyderabad, 13-15 December, 2002.
Web : <http://www.languageTechnologies.ac.in/lec2002.html>
- International Conference on online learning, NCST, Mumbai, 15-17 December, 2002.
Web : www.ncst.ernet.in/vidyakash2002
- ICON (International Conference on NLP) 2002, & International Conference on Knowledge Based Computer Systems, NCST Mumbai, 18-21, Dec. 2002.
Web : <http://www.ncst.org>
- A Workshop on NLP and Applications, Anna University, Chennai, Dec. 10-21, 2002.
Web : <http://annauni.edu>
- Workshop on Spoken Language Processing, Tata Institute of Fundamental Research, Mumbai, Jan. 9-11, 2003.
Web : <http://speech.tifr.res.in>
- Indo Wordnet Workshop, CIIL Mysore, Jan. 14-15, 2003.
Web : <http://www.ciil.org>
- Second International Workshop on Development in Indian Languages (IWTDIL-2003), ISI Kolkata, Jan. 22-24, 2003.
Web : www.isical.ac.in/~cvpr/events
- 13th international workshop on Research issues on Data Engineering: Multilingual Information Management (RIDE-MLIM' 2003), March 10-11, 2003.
Web : <http://www.aztecsoft.com>
- JHU Summer Workshop on Language Engineering, Baltimore, Maryland, USA, from July 14 to Aug. 22, 2003.
Web : <http://www.clsp.jhu.edu/workshops>

Unicode Related Meetings & Events

- UTC#92/L2#189 hosted by Microsoft, Redmond, WA, August 20-23, 2002.
- 22nd International Unicode Conference, San Jose, CA, Sept. 9-13, 2002.
- SC22/WG20 Japan (Tentative), October 2002.
- UTC#93/L2#190 Annual Members Meeting hosted by Compaq, Nashua, NH, November 5-8, 2002.
- IRG # 20 Hanoi, Vietnam, November 18-22, 2002.
- SC2/WG2#43 Tokyo, Japan Dec. 9-13, 2002.
- UTC/L2/IUC Mountain View, CA January 31, 2003.
- SC22/WG20 Busan, Korea (Tentative) Feb. 11-13, 2003.
- UTC#94/L2#191 Mountain View, CA March 4, 5-7, 2003.
- 23rd International Unicode Conference Prague, Czech Republic March 23-28, 2003.
- IRG#21 China April 7-11, 2003.
Web : <http://www.unicode.org/unicode/timesens/calendar.html>
Note : ■ Indicates conferences in India.

2. Reader's Feedback

“Global digital divide among scripts”

*Thanks for including me in your mailing list.
Attached is my short commentary on the theme of VishwaBharat.
I hope we are sharing the same interest.*

Recently I visited a website of the United Nations. The site (<http://www.unhchr.ch/udhr/>) introduces more than three hundred different language versions –from Abkhaz to Zulu!– of the “ Universal Declaration of Human Rights”. The site claims that this text is the most widely translated text in the world, and has been awarded the Guinness World Record for having done this great job. Thus the Universal Declaration is indeed “ the most universal text” in the world.

Try now! and you can find all eighteen Indian official language versions of the 1,778 words text, with only two exceptions – Konkani and Manipuri, But really disappointing for you would be the fact that all-Indian language versions are just posted as “gif” files, not in the form of encoded texts. And actually many other non-Latin scripts users in the world have to feel the same kind of sadness after visiting. Please note that I am not blaming staff of the United Nations at all, rather highly appreciating the great efforts of the staff and many voluntary supporters involved to create such valuable web resources. What I again recognized is the most urgent need of creating “equal opportunities for every language and script users around the globe”.

My recent study based on statistical data provided by the International Telecommunication Union and UNESCO gives a rough sketch of global digital-divide “among script”, Latin alphabet users, 39% of global population, consume 72% of world total writing/printing paper consumption and enjoy 84% of access to the internet. Hanzi-Chinese Ideograph –users in China/Japan/Korea, 22% in global population consumes 23% of paper and have 13% of internet access. Arabic script users, 9% in population consume 0.5% of paper and have 1.2% of Internet access. Cyril script users 5% in population, consume 1.1% of paper and have 1.6% of internet access. Then how about Indic script users? If all Bralmi-origin scripts widely used in southeast Asia-Myanmar, Thai, Lao, Khmer, etc.-included, Indic scripts users occupy 22% of world population, consume 2.2% of paper, and have just 0.3 of internet access.

The University Declaration of Human Rights Article 2 says: “Everyone is entitled to all the right and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language (underline), religion, political or other opinion, national or social origin, property, birth or other status. **TDIL is really needed!**”

*Yoshiki Mikami
Professor, Management and Information Systems Engineering
Nagaoka University of Technology, Japan.
E-mail : mikami@kjs.nagaokaut.ac.jp
Fax:- 81-258-47-9350*



Need for Sanskrit Software in Public Domain

...Congratulations to your successful activities. You have done a lot! ...I have "borrowed" a font, arranged by copy and paste within a font editor, but it would be better if I can get a free one. Do you also have the possibility to ask someone to design a new font, based on an old one?

I installed Desika under WinME and according to its description I expected an elephant, but what I found was an ant, which only demonstrated subanta. This is not the way how software people all over the world handle demos. For casual users it would also be fine to explain that the fonts have to be installed.

My motivation for doing something in this field is that Sanskrit language, its literature and its techniques (science of life) have to be maintained, because only this knowledge can help this world of today, where the life values are decaying. The "complicated" Panini grammar obviously was the result of making the spoken word powerful (Mantras, rituals etc.). The Turkey language e.g. has similar structure, but almost no exceptions. The sound effects were not known.

I found that many things which I plan to do are already done in the ILP Tools or are to be done. The question is, how much of them is available for me?

Here are some things which I would like to have:

- 1) I have a list of Sanskrit words, which can be used for syntax check, but unfortunately with spelling errors and a not unique accent representation. Here directly from the text file which is in the Harvard- Kyoto transliteration (it looks odd):

a4—tIkSNa, a4-zuci or a-zuci4, a4Mza—karaNa

The number 4 represents both udaatta and swarita. Are there rules that help to find out, if it (number 4) represents either udaatta or swarita, e.g. position in the word, preceding or following letter etc.? Otherwise all this kind of words have to be checked manually.

- 2) Which rules are available to find out the correct spelling of Sanskrit words? We want to use them for Syntax Check.
- 3) I have collected all conjuncts (consonant compounds) from different books. Are there rules, which allow or don't allow certain combinations? In other words, I need a complete list of all possible conjuncts, or the corresponding rules, which automatically excludes those consonant combinations, which are not allowed. This is also for syntax check.
- 4) I am interested in the procedures and rules to resolve samaasa.

- 5) Furthermore it would be very helpful to have a collection of all endings of words with their syntactical and semantical meaning.

- 6) Panini Rules in text form for faster search (instead of searching in a book) or in an implemented form for analysis.

Here are the ILP Tools, which cover a lot. I would like to have their source code and detailed documentation so that I can implement them in my system. Important ones are : Desika, Sanskrit Authoring System, Syntactic and Semantic Analysis of Sanskrit Sentences, Computer Assisted Sanskrit Teaching & Learning Environment, Shabdhabodha.

Proposal

Please forgive me, if I am making the following proposal to you, probably out of ignorance of Indian situations, but I cannot resist to do so.

First of all I think, results of projects related to Sanskrit and sponsored by the Indian government, should be open to all (open source, open documentation), so that the progress can be enhanced. Suppose, one developer leaves the Institute, his work might be lost, and also the money which was spent on him. There are not many programmers in this field!

A central coordinator may collect every 2 -3 months the results and check the progress. Then not much can be lost, and the developer is checked.

Other developer may use one or the other component, i.e. source code or data, and by this he accelerates his own work. And double work is avoided- which is obviously there, documented in the description of the ILP Tools.

"Trust is good, control is better" (Lenin).

There should be competition, parallel work can also be done, because there might be different ideas and sometimes it cannot be decided which one is better. But the competition should be in a positive way, so everyone helps the other by showing his results. It is also no harm if one party fails on its chosen way, because other parties will not go the same way and by this save time.

Whatever help I can offer, I will do.

With warm regards,

(Prof. Gerd Unruh
Cornelia Schlosser Allee 5, D- 79111 Freiburg, Germany
Tel: 0049 761 4880 335, Fax: 0049 761 4880 337
E-mail : un@fh-furtwangen.de)