COMPUTATIONAL ANALYSIS OF SANSKRIT LANGUAGE

Anirban Dash, Ph.D.

COMPUTATIONAL ANALYSIS OF SANSKRIT LANGUAGE
Anirban Dash, Ph.D.

1. SANSKRIT FOR COMPUTER

In July 1987, *Forbes* magazine published news, which surprised even the Sanskrit pandits. It said that, "Sanskrit is the most convenient language for computer software programming." It filled the hearts of all those who love and study Sanskrit with great joy and enthusiasm as it opened the doors to new fascinating world of Sanskrit studies.

Sanskrit is extremely rich, powerful and expressive language. Its potentialities are gradually being appreciated all over the world, and its application is being extended in different fields.

To mention a few, Indian Institute of Information Technology (IIIT), Hyderabad is working on a project called Natural Language Processing (NLP). The goal of NLP is to build computational methods of natural language for its analysis and generation. Similar work is going on at C-DAC, Bangalore. The project is called Natural Language Understanding (NLU).

2. COMPUTER FOR SANSKRIT

The association of Sanskrit with computer is constantly bridging the distance between the science of ancient world and the world of modern science.

Computer scientists and Sanskrit pandits all over the world are trying to use the computer technology in the field of Sanskrit studies. Many such efforts have already started in India. In this regard, Government of India has initiated two major projects:

1)   Technology Development for Indian Languages (TDIL)
2)   Sansk-net project. (site name: **http://www.sansknet.org**)

3.   TECHNOLOGY DEVELOPMENT FOR INDIAN LANGUAGES (TDIL)

The Indian Standard Code for Information Interchange (ISCII) was devised by C-DAC, for using any Indian Language in Word Processing, Data Processing and a host of other applications across platforms like DOS, WINDOWS, UNIX, Mac, etc. This standard provides instant transliteration of texts among Indian languages and Roman script (with diacritics).

4. SANSK– NET PROJECT

The Project Sansk-net was proposed by the Indian Heritage Group (IHG) and Real-Time Systems Group (RTSG), and Center for Development of Advanced Computing (C-DAC), Bangalore, to be an initiative with Rashtriya Sanskrit Vidya Peetha (RSVP), Tirupati as

the nodal center and C-DAC as the technical implementation agency. The scope and objectives of the project are as follows:

Objectives

The basic objectives of the "Sansknet" program are the following:

1. To present the database available in different institutions in a Computer framework.

2. To assist the institutions to develop the hardware, software and the technical capability to place the information in the modern technical framework.

3. To develop computerized linkage among the different institution so that each institution can have access to the database available in the other institutions.

4.To make use of the principles and techniques available in *Nyaaya, VyaakaraNa, Vedaanta* and *VedaaNga* for developing new paradigms for the computer.

5. To develop packages for training for the faculties in the scientific work and *shaastric* world for making best use of the infra-structural facility.

6. To facilitate preservation of the information on rare manuscript, *Vedic* literature and *shaastras.*

5. COMPUTER ORIENTED SANSKRIT STUDIES

At present, computer-oriented Sanskrit studies are heading in following directions.

Development of corpora of Sanskrit Text

Corpora, generally, refers to the complete collection of writings in a machine-readable form, which have different uses in linguistics and its application, such as: preparation of lexicon, grammar, dictionaries, thesaurus, frequency list and the like. It would not only provide basic research facilities for studying linguistic features but also meet the practical need of translation, knowledge-based compilations and such other applications. Many institutions and persons have engaged themselves in the development of such corpora. Some such sites have already gone online. Indian Heritage group under the Ministry of Human Resource Development (HRD) is building a corpora of ancient Sanskrit texts related to different areas like: *Veda, Upanishad, PuraNa*, Epics etc.

Some institutions are working to build Corpora on various ancient texts under Indian Heritage Group (IHG). The names of the Institutions are as follows:

1)   Chinmaya Foundation (CF), Cochin.

2)   PuraaNa PrajñÁ Suddhodhana Mandiram (PPSM), Bangalore.

3) Academy of Sanskrit Research, Melkote (ASRM).
4) The Rashstriya Sanskrit Vidyapeetha (RSVP), Tirupati.
5) Ahobila Math Sanskrit College (AMSC), Madhurantakam.

As a matter of fact, the following texts have been entered by these Institutions.

Name of the Institutions	Name of the some texts	Number of the entered texts
Chinmaya Foundation, Cochin	Tantravaarttika BharadvaajagRhya SUtra Upanishad–ShaaNkara BhaaSya IÐÁvÁÒya UpaniÒad Kena UpaniÒad ÏÁbarabhÁÒya etc.	33
PuraaNa PrajñÁ Suddhodhana Mandiram (PPSM), Bangalore.	SUtra Deepikaa Tattva MaNjaree Bhaava Deepikaa Parabrahma Prakaashikaa Nyaaya VivraNaÞ etc.	27
Academy of Sanskrit Research Melkote. (ASRM).	RigVeda Yajur Veda SÁmaVeda AtharvaVeda VedaaNgas etc.	33
The Rashtriya Sanskrit Vidyapeetha (RSVP), Tirupati. Ahobila Math Sanskrit College (AMSC), Madhurantkam	DhvanyÁloka KuvalayÁnanda with Chandrika Dharmavijaya NaaTaka RukminipÁnigrahaÆa etc. VakroktijeevitaÞ KÁvyameemaaÞsaa Ïri BhaaÒyaÞ Vedaanta Deepa Vedaanta Saara etc.	78 24

6. MACHINE-AIDED TRANSLATION

Machine Translation means the translation with the help of machine. It is an effective transfer of textual materials from one language to another language with the help of computer. This involves dictionary search, selection of equivalent terms, morphological information and error corrections, etc. Department of Computer Science and Engineering, IIT, Kanpur, from 1983 onwards, has undertaken a project attempting to utilize the Sanskrit grammar structure and particularly, PaaNini's theory, for an interlingual-based machine translation system among pairs of Indian languages. This project has achieved considerable success by now, covering Hindi, Telugu, Kannada, Marathi, etc.

7. E-LEARNING PROGRAM FOR SANSKRIT

Sanskrit is taught from primary level to University level in different parts of the country in both traditional and general institutions. It is also learned in many foreign countries. However, the process of learning and the teaching is not up to the mark for various reasons. Besides, there is no facility for learning Sanskrit independently. Thus, there is a need to develop linguistic technique aided by audio-visual interactive media, which can take care of requirements of various learning groups. The work in this direction is already on in India and abroad. Sanskrita Bhaaratee and Vishva Bhaaratee have already prepared an E-learning program for basic Sanskrit. Similar work is also going on at the University of Pune, Department of Sanskrit in collaboration with C-DAC multimedia group. This E-learning and Expression project aims at preparing programs for different levels in addition to *SambhaasaNa* portal, i.e., package for spoken Sanskrit.

8. COMPUTER-ASSISTED SHAASTRIC TEACHING

Sanskrit *Shastras* are usually taught at higher level. The study of such *shaastras* includes the study of main text along with its commentaries and sub commentaries. When one studies a particular *shaastra*, it is necessary to understand the ideas expounded in that particular *shastra* in their proper prospective, for which one is required to go through a number of books, commentaries and dictionaries in order to collect all the scattered details. Without the help of modern technology it would be difficult to get all the relevant information, scattered over many texts or parts of single text. Computer can be used as an effective means of inputting, encoding and scanning the data. It can retrieve the desired information in a flash and can present it in a systematic way. We can thus develop a larger system of information, accessible through electronic network, to provide a better textual exposition.

Some such attempts have already been made. To mention a few, there is a package called 'Gita Super' developed by IIT, Kanpur, which presents text of *Bhagavatagnta* along with its two commentaries. Another package called the *MahÁbhÁrata* is developed by Bhandarakara Oriental Research Institute (BORI), Pune. Similar works are carried out abroad too.

9. COMPUTATIONAL ANALYSIS OF SANSKRIT

Computational analysis signifies an analysis of given text with the help of computational techniques. The computational analysis of a Sanskrit text includes a number of procedures such as:

1) Keying the text in any editing software.
2) Analysing the text from syntactic and morphological point of view.
3) Preparing a detailed help program for grammatical analysis.
4) Developing programs for sorting, searching, and indexing, preparing concordance, creating hyperlinks for words, rules and verses.
5) Creating database for storing all the information.

Though, a variety of Natural Language interface tools like an Editor, creation of multi-lingual documents with transliteration between the scripts of Indian languages and the Roman script, utilities for sorting, searching, indexing, concordance, various analyses like morphological, syntactic and semantic, lexical update, grammar help, hyperlinks to a variety of rule bases, etc. are developed for this purpose.

Their effective application for producing better results is still a great challenge. Computational analysis of Sanskrit Sha*stras* poses still greater challenge as it not only demands the knowledge of computer technology but also a thorough knowledge of the concerned text.

A preliminary Natural Language Understanding (NLU) System for Sanskrit has been developed and is in use in universities and institutions.

Essentially, it generates/analyses Sanskrit Words/Sentences with the help of PÁnÍnian grammar rules, for valid word forms. Paraphrasing an input sentence, changing the voice, euphonic combinations and accented input (*Vedic*) processing are the other salient features of this system. Rule base trace is also provided.

10. TO CONCLUDE

The advanced computer technology has revolutionized almost all parts of human life and field of education is no exception.

This attempt to use the modern technology for analyzing the ancient works will be helpful to the scholastic community as a whole in learning, understanding and interpreting ancient Indian knowledge in a better way.

Such packages can also prove useful in teaching ancient Indian *shastric* texts effectively and with more efficiency. It can also be used as supporting systems for research. Thus, there is an immense potentiality in this field, which needs to be explored further for a better result.

In this way we can develop search packages based ancient texts. It is quite obvious that, for preparing such packages, the knowledge of Computer and knowledge of Sanskrit should go hand in hand and complement each other for better results. What is required in future is the increased co–operation between a Computer professionals and Sanskrit Pandits.

Anirban Dash, Ph.D.
Central Institute of Higher Tibetan Studies
Sarnath-221007, U.P.
India
srianirban@yahoo.com