UTILITY OF COMPUTERS IN VEDIC KNOWLEDGE-BASE CREATION

By *VedaVaridhi* **Dr. P. Ramanujan**

Programme Co-ordinator, Indian Heritage Group, C-DAC, Bangalore (KP)

Abstract

We describe the way modern advancements in the field of Computing can be put to good use in creating Vedic knowledgebase for preservation, study, analysis, modeling and propagation of Sanatana Dharma through fourteen Vidya-sthanas.

Report of work done at IHG, C-DAC, Bangalore, since 1990 and future plans are also articulated for eliciting co-operation and collaboration from traditional domain experts of Vedic knowledge.

A case study is also presented to illustrate the idea in practical terms. A demonstration of the systems developed is also attempted.

Introduction

Computers pervade every conceivable human endeavour in today's world and it is wise to harness them for our benefit in situations where speed and memory are important; complex, repetitive, time-consuming, monotonous jobs like storing huge data, searching through them, finding patterns, indexing, cross-referencing, hyper-linking, transliteration, analysis etc. are best done using Computers. The field of Artificial Intelligence and Natural Language Processing along with Pattern Recognition, Logic Programming etc. help Semantic and Pragmatic Analyses.

IHG, C-DAC has contributed to such efforts by way of evolving PC-ISCII standards, Computational Rendering of Panini's grammar, developing NLU System DESIKA, Sanskrit Authoring System C-VYASA, with database, Digital Library of Indian Heritage, including Vedic and Vedanga texts, Ramayana, Mahabharata etc. and also pioneered the development of tools for carrying out the type of jobs enumerated above. We are currently engaged in creating a Heritage portal encompassing these aspects. Utilities for Palm-leaf manuscript study are also under development.

Vedic knowledgebase

There are fourteen vidya-sthŒnas or dharma-sthŒnas, covering four Vedas, six VedŒngas and four UpŒngas as enumerated by Sages VyŒsa and YŒjnavalkya. "*PurŒÄa-NyŒya-M¶mŒÃsŒ Dharma-ÝŒstrŒÆga miÜritŒ¬. VedŒ¬ sthŒnŒni vidyŒyŒ dharmasya ca caturdaÜa ..".*

Here, the digital web contents are created for all the fourteen vidyasthana-s, viz. *Veda-s* (scriptures - 4): RgVeda, YajurVeda (Sukla & Krsna), SamaVeda, AtharvaVeda; *Vedanga-s* (Vedic auxiliary sciences - 6): Shiksha, Vyakarnam, Chandas, Niruktam, Jyautisham, Kalpa; *Upanga-s* (supplementary sciences/subjects - 4): Mimamsa (Including Vedanta), Nyaya, Puranas, Dharma Shastra.

The a´gas and upı´gas of the Vedas are to be used for the proper interpretation of the Vedic texts, which alone are the sole repository of all knowledge leading to various attainments (material and spiritual).

These also have four phases each, viz, Learning, Reflecting, Practice and Propagation.

Vedic Processing Tools

The Vedic Editor and Sanskrit Authoring System, *C-VYASA*, developed by, us has a basic Editor which provides for data entry including accent-marking in Devanagari and Grantha scripts, transliteration between them, find, replace, pattern search (even accents separately), phrase search (which includes accented and unaccented, script dependent and independent options, to give a concordance listing in which all texts the search phrase occurs, how many times in each and where, along with a provision for saving the search results to a file), printing, copy, paste, convert between different formats and to database format, create word index, sort in several orders etc.

The knowledgebase of the fourteen vidyasthanas at the current level, is also included for insertion of any quotation from standard texts of Vedas/shastras, in an article a scholar might be writing (typing), by searching in the relevant application program appropriately and copy, pasting to the editor with location reference. The word index of the article and quotations can be generated. Among Vedic texts, this includes RgVeda and YajurVeda currently and SamaVeda is separately provided. About 20 different application programs are included. A lot more texts need to be added to the knowledgebase to have an authentic, comprehensive reference compendium of original treatises, but the job is on. Presently, about 270 texts covering all the vidyasthanas are available. Global copy feature and exporting to MS-WORD, RTF or html formats are provided. Linking from Vedangas to Vedas is another feature between Shiksha, Nirukta, Chandas, Kalpa, Vyakarana and Mimamsa, as of now.

Pandu-lipi Samshodhaka - A Computer-based Manuscript Editor

Introduction

Facilities to preserve, study and publish information contained in manuscripts (palm leaf, paper, etc.) with the help of advanced computing tools and technologies are welcome in order to unearth the treasures hidden in them for betterment of mankind.

C-DAC has developed a comprehensive Manuscript Processing Software ***Pandu-Lipi Samshodhaka*** for the purpose. We describe the salient features of this further.

Background

Having evolved PC-ISCII standards for proper representation of all Sanskrit and Vedic character set in computers, developed an exhaustive knowledgebase of Vedas, Vedangas and Upangas and application programs for the fourteen Vidyasthanas and tools and utilities like Editor, index, search, concordance etc., we now undertake to extend these to the deciphering

of manuscripts in scripts like Grantha, Nandinagari, Telugu, Malayalam etc., of Sanskrit/Vedic texts, many of which are not yet published.

We also help collation of various version/variant forms of texts from different sources for critical editions of such rare, unpublished works in these domains.

Features

The functional modules of the system cover Acquisition, formatting, inputting, indexing, creating database, searching, locating, printing, collation and publishing.

The range of texts covered include shastric texts, RgVeda, Krishna Yajurveda, Samaveda, Lakshana Granthas, texts in Tamil, a combination of Tamil and Sanskrit, called Manipravala, etc., in a variety of scripts. The sample includes about 50 works, 120 manuscripts, 6 scripts and many domains. There are about 3500 leaves (pages) as images to accompany the PC-ISCII texts.

The text Shadvimsati Sutra is chosen, for illustration and publication of a critical edition with a Sanskrit commentary by the author.

Description of the modules

Acquisition

One typically starts with consulting catalogues, indices, lists, reports, etc., of Manuscript collection of desired texts through a number of sources Bibliographic survey. Shortlisting the ones feasible to obtain from the list, providing for balanced representation of various regions, script and versions (i.e. with commentaries, with accents etc.) and acquiring copies (Xeroxed/scanned /microfilmed), converting/exporting to a single (uniform) format i.e., jpg in the current case are the further steps. Factors like clarity, condition of original, resolution of scanning and size of the image files influence the choice of the common format.

Formatting

The inputs come in variegated forms, when raw, i.e., from the institutions/library collections. Even cost considerations (since usually copies of Manuscript are charged on per-exposure basis) and good-quality multi-folio image files while scanning are perfectly possible. We may have 3,5 or even 10 folios per scanned image. Here, the two sides of the folios are in different files and the job of sequencing the image as per text and separation of folios is involved. Numbering them serially according to the text is done.

An important task here, in the case of Manuscript bundles containing different texts, is separation of the texts and folios belonging to multiple texts. They must be present in all the works concerned. Usually, libraries offer separation, if catalogued already.

Inputting

We strongly recommend the entry of the data contained in the manuscripts for the purpose of study, word-split, index, search (phrases), editing and collation. This, of course, requires domain experts who are difficult to get. However, IHG can offer expertise in this endeavor.

We also have another possible source for data entry, which is loading text, if the work in the manuscript is one of available digital texts from our repository (a list of about 250 texts

from all Vidyasthanas is available. C-DAC Indian Heritage Portal would make this available on the web soon).

Adding commentaries, translations, hyperlinks, annotations for collation etc, are the factors necessitating data-entry. Also transliteration, training in rare scripts etc, is enabled. However, efforts may be launched to develop efficient OCR or speech recognition systems of high quality simultaneously and when these mature, we can minimize data entry needed.

Editing

This step involves aligning the data entered, with the original manuscript, line by line and page by page. This also can be done in an editbox/window below (or adjacent to) the image of the manuscript or entered through Vedic Editor and inserted into database. The pages and line boundaries are as before. Adding information for retrieval, hyperlinks etc., can also be done.

Multilingual texts, currently require LEAP-like software for data entry and use in RTF format in the system for further processing. Here ISCII-ISFOC conversion is employed.

Currently Vedic texts of Samaveda Gana require use of only Grantha script and transliteration is not available. Srautam and Guruparampara Prabhava etc. are multilingual samples. These are typed in LEAP and processed through rtf controls.

Creating database

The PC-ISCII text files (*.pci) created by data entry or loading data are to be converted into database format. This is either Microsoft Access or Microsoft FoxPro format covering various fields for facilitating information retrieval. There is a utility that converts from aci/pci format to db format.

Databases of works, institutions, manuscripts, books etc., are also created and linked in the application list of abbreviations. Scheme of data for reference in these texts etc. are also created as tables.

Searching

This is the crux of the system and helps in providing word or phrase level search (with and without accent-markers) across the database, text-wise, and lists the Manuscript ref. nos. where the search string occurs. In future, we can even extend this across texts if need be (this feature is there in our Vedic Editor, wherein a string occurring in any of the 250+ texts are listed as a concordance).

Choice of script, facility to transliterate and seeing the results in the same manner of alignment as the manuscript are the useful aspects.

Locating

This refers to locating the search string in the image of the particular page of the manuscript where it occurs, including the line number and location in it. We see the string ‘highlighted’ in the text window by choosing ‘find’ in the page and physically looking in the corresponding line and ‘location’ in it on the image above by selecting view in ‘search’ mode. The text window is provided with line numbers to facilitate this manual locating in the image.

Printing

Provision to print the texts in database, search results etc., in any script of choice or script of the original etc. so that further reference or insertion into documents can be enabled. Report generation kind of printing needs can also be addressed. List of texts, institutions, reference details etc. can be printed.

Collation

From the search function, we can organise the readings of different texts (like 'file compare') across the manuscripts combined with report generators. A scheme for annotating can be devised to assist here. Work will follow to enrich features here.

Publishing

Publication through Desk-top-publishing can be done by exporting to some DTP software and adding embellishments as desired.

A Case Study of Shadvimsati Sutra

We now present the findings of a *Case-Study of an unpublished, rare, important Vedic text being brought out as a critical edition using the Computer software Pandu-lipi Samshodhaka,* as an illustration. The text released in this session is the one produced so. An English translation is being added to this work to reach out to more people.

Name of the Text:	**Shadvimsati Sutra**
Belongs to Veda:	Krishna Yajur Veda
Type of text:	Lakshana-grantha
Brief Description:	Ref. No. 1335 of Dr. Parameshwar Aithal's Vedic texts Index - 1335 SADVIMSATI(-SUTRA) - This work is intended to help learners of the Taittireeya-Samhita by removing the confusion arising from pairs of similar words, such as "prajapati" and "prajavati".

List of available sources:

A. Manuscripts

1. Adyar D I 1029 (25 L 35 = 71153). PL. Gr. 5 fol. (18ff.).
2. Adyar D I 983 -> Veda-lakshana (A 1).
3. Adyar D I 984 -> Veda-lakshana (A 2).
4. Adyar D I 985 -> Veda-lakshana (A 3).
5. Adyar D I 986 -> Veda-lakshana (A 4).
6. Adyar D I 989 -> Veda-lakshana (A 7).
7. *Adyar D XIII 195 (VB 115c). PL. Tel. 6 fol. Incomplete.
8. Adyar D XIII 196 (VB 461k). PL. Tel. 9 fol. Incomplete.
9. *Adyar D XIII 197 (VB 1141d). PL. Na. 13 fol.
10. *Baroda 2567. 11 fol. 150 Grantha-s. = CLB I, p. 37.
11. Baroda 10034(h). PL. Gr. 4 fol. (17-20). 120 Grantha-s. Incomplete. = CLB I, p. 37.
12. Baroda 10381(a). PL. Gr. 12 fol. (5-16). 150 Grantha-s. Incomplete. = CLB I, p. 38.
13. BrI. P. 37 (no.CXX) = IO 4510.
14. Hamburg StUB (cod.palmbl. III 8/133) -> Veda-lakshana (HB).

15. Hamburg StUB (Oppert 10d). PL. Gr.
16. Hoshiarpur p. 54 (no.2025). PL. Gr. 47 fol. 700 Grantha-s. Extract p. 413 (the extract printed here is that of Phulla-sutra, ch. 3!).
17. *Hoshiarpur p. 54 (no.4375) PL. Gr. 8 fol. 200 Grantha-s. Extract p. 413.
18. *Hoshiarpur p. 54 (no.5040) 8 fol. 118 Grantha-s. Extract pp. 413-14.
19. Hz. 627. PL. Tel. 24 fol. (with 4 other works). Belonged to Vedam Venkatasubrahmanya Somayaji of Allur, Nellur Dist.
20. IM 2049. 12 fol.
21. IO 4510 (Burnell 146h). PL. Gr. 8 fol.
22. IO 4511 (Mackenzie III 59b). PL. Tel. 8 fol.
23. IO 4510 (Mackenzie III 186d). PL. Tel. 9 fol. (5-13). Written ca. A.D. 1800.
24. *MD 991. PL. Tel. 23 pp. (22ff.).
25. *MD 992. PL. Tel. 15 pp. (81ff.).
26. MD 993. PL. Na. 4 pp. (12ff.). Incomplete.
27. MD 994. PL. Gr. 53 pp. (39ff.). Incomplete.
28. MD 995. PL. Gr. 12 pp. (7ff.). Incomplete.
29. MD 16706. PL. Tel. 6 pp. 96 Grantha-s.
30. MD 17447. PL. Gr. 20 pp. 210 Grantha-s.
31. MT 485(q). PL. Tel. 8 fol. (70-77). Presented in 1910-11 by C. Visvanatha Shastrigal of Viziyanagaram.
32. MT 1964(j). PL. Tel. 6 fol. (99-104). Presented in 1916-17 by Rayaprolu Ramabrahmangaru of Guruza, Gudivada Taluk, Krishna Didt.
33. MT 3887(k). PL. Gr. 4 fol. (85-88). Purchased in 1921-22 from Annasvami Sastrigal, Bhavani, Coimbatore Dist.
34. Mysore I, pp. 36-37 (10 Mss.).
35. Mysore ND II 3470 (P 4626/ 2). PL. Tel. 19 fol. (59-77). Extract no. 167. Title: Taittireeya-samhita-varnanukramanika.
36. Mysore ND II 3641 (P 3280/ 3). PL. Tel. 9 fol. Extract no. 226. Title: Pluta-nirnaya.
37. Mysore ND II 3851-63. Extract no. 293.
38. Oppert I 2473. PL. Gr. 42 pp. Belonged to Sri Sarasvati Bhandaram Committee of Tiruvallikkeni, Madras.
39. Oppert I 7248. Belonged to Ivattur Nandikesvara Sastri of Vijayanagaram.
40. Oppert II 782. Belonged to Ghanapathi Shingaracharyar of Alasur, Madurantakam Taluk.
41. Oppert II 1393. Belonged to Gopalacharyar of Pillappakkam, Kanchipuram Taluk.
42. Oppert II 1530. Belonged to Sri Shankaracharya Matha of Sringeri.
43. Oppert II 7989. Belonged to Venkatarama Dikshita of Sorakkayur, Thanjavur Dist.
44. PUL I, p. 22 (no.101). PL. Gr. 56 fol. (75-130).
45. PUL II. App. P. 6 (no.99). PL. Gr. 10 fol. (121-30).
46. *RASB II 487-88(VIII) (G 6063). 10 fol. (18-27). V.S. 1913 / Saka 1778.
47. *SBD 4159. 14 fol.
48. SBD 55308. 5 fol. (sadvimsati-laksana).
49. *Tirupati SVUORI 7052 (4421b). PL. Gr. 13 fol.
50. Tirupati SVUORI 7053 (900a). PL. Gr.
51. *Trav. Uni. 2346 M. PL. Gr. 150 Grantha-s.
52. *Trav. Uni. 2388 A. PL. Gr. 150 Grantha-s. Damaged.
53. *Trav. Uni. 2938 K. PL. Gr. 150 Grantha-s.
54. *Trav. Uni. 5512 B. PL. Gr. 150 Grantha-s.
55. *Trav. Uni. 16408 U. PL. Mal. 9 fol. 150 Grantha-s.

56. *Trav. Uni. L.988 M. PL. Mal. 7 fol. 140 Grantha-s. Incomplete.
57. *Ujjain I, p.10 (no.1118). 7 fol.
58. *Wai II 10264. 9 fol.
59. *Wai II 10265. 8 fol. Dated Saka 1730.
60. *Wai II 10266. 9 fol. Dated Saka 1739.
61. Wai II 10267. 14 fol.
62. *Wai II 10268. 12 fol. Dated Saka 1753.

B. Printed Books

in Grantha from Kumbakonam by Vaidyanatha Shastri in 1899 1st Ed., 1922 2nd Ed.; in Telugu from Mysore by Gomatham Srinivasa Jyosyar in 1919.
[These two are available in my personal collection and both have been scanned].

C. Commentaries

C. Oppert II 783. Belonged to Ghanapathi Shingaracharyar of Alasur, Madurantakam Taluk. [CC I 679b; II 162b]

Manuscripts marked with asterisk are used in the current critical edition. Additionally, a manuscript from Deccan College, #9387, was also used. From ORI, Mysore, out of the first in the list, i.e, item 34 above, two Mss. #144 and #2262 were used. Both the printed versions are considered. In all, 28 manuscripts have been consulted.

What needs to be done? Plans for the domain

- Standardisation of Accent-marking scheme for all Vedas and phonetic correlation data
- Introduction of Salakshana, Sanga and Bhashya studies in Vedic institutions
- Vedic content creation using advanced technologies for research and publication of rare texts from manuscripts
- Training for traditional scholars and familiarization of shastraic concepts for modern youth, specifically, linguistic and Computer Science streams in interdisciplinary research
- Creation of reference works like indices, concordances and treatises like dictionaries for comparative study, using the tools developed
- Translation into regional languages for better reach and understanding, including multi-media productions

Conclusion

We need more interactions in the form of seminars, workshops, conferences, Vidvat Sadas (such as the present one), monographs, hands-on sessions, specialized training to keep the efforts continuing and build from previous efforts collectively, to benefit the entire Universe. Revision of academic curricula in Language and Computer Science streams is a desideratum. Training of teachers and scholars to be useful as researchers, resource persons and project guides is also to be planned and implemented. All future publications of Vedic texts

may use these tools and have value-added features like indices, reference, bibliography, hyper-links in e-book format etc. They could be accented where necessary.