ORIENTALIA PRAGENSIA

INDOLOGY IN E-SPACE

PETR DUDA, CHARLES UNIVERSITY IN PRAGUE

The paper attempts to outline the current position of Indian studies in the world of the Internet, introducing the most remarkable Indological online projects, categorizing the available tools and resources and considering the possible directions that future development could take.

The topic of the present paper is of rather marginal interest for academic Indology, yet I believe that it has its legitimate place in the present discourse on the search for Indological identities. It is perhaps not too exaggerating to say that when visiting the websites of many departments of Indology today, we will come across an increasing number of tools and resources for – mostly linguistic – study and research, rather than just a bibliography of their members or a list of the subjects taught or researched. There is one more area that could be occupied by Indology on the web, namely presenting basic information about Indian culture, history, literature etc., intended for the general public and comparable to the popularisation literature which has always been one of the traditional roles of at least European Indology. It has been mentioned many times that the general notions about India are often surprisingly mistaken in Europe. It is questionable, however, whether Indology should and will bring the exercise of this role to the Internet, where it has to compete with and to correct the many misinterpretations in the online presentations of various religious movements and in enthusiasts' websites, or whether this ambition is to be given up and the space relinquished to such alternative views.

In this modest contribution to a most interesting debate, we will focus on the first of the above-mentioned forms of presence of Indology on the Internet, that is, the online tools and resources intended for students, teachers and researchers in the field of Indology, or to be more precise, classical Indology in its nowadays questioned sense of the word, i.e. the study of Indian languages and especially Sanskrit. Once again, we can see a pattern of the most attention being given, at least initially, to Sanskrit language and literature. It has to be stated that we are not going to deal with computational linguistics as such, even though some of the results of this highly interesting field will be introduced. Neither is it our ambition to cover all areas and to introduce all the projects that would actually deserve to be covered. I would like to draw the reader's attention to just a few selected projects or websites that will illustrate the development trends and the potential available for Indian studies in the e-world. The below-mentioned projects have been developed by academics at – mostly but not solely – Indological departments in India, Europe and Northern America.

The presence of Indology in e-space naturally reflects also the limitations and achievements of information technologies in general. Thus, the format of the transcription and encoding of diacritics and its satisfactory display in web browsers had to be dealt with first. This was followed by the process of digitalizing a large part of the most important Sanskrit works. Meanwhile, reliable dictionaries of Sanskrit as well as those of modern Indian languages were converted online. Finally, the results of automated analysis of the Sanskrit language, or even Sanskrit to Hindi machine translation projects started to appear. The three main categories of online tools currently available for the study of Indian languages, and particularly of Sanskrit, thus include dictionaries, electronic text corpora and Sanskrit grammar applications.

It is probably unnecessary to introduce the Cologne Digital Sanskrit Dictionaries Project of the Institute of Indology and Tamil Studies at the University of Cologne.[1] In addition to the digitalised Monier Williams' *Sanskrit-English Dictionary*, Böthlingk's and Roth's *Kleines* as well as *Grosses Petersburger Wörterbuch* and Apte's *English-Sanskrit Dictionary*, there are now 4 more scanned dictionaries available plus a scanned Kale's *Higher Sanskrit Grammar*. Furthermore, the website offers an application for the automated recognition of inflected forms, displaying the corresponding dictionary entry word and the entire paradigm. Numerous versions of the digitalised Monier Williams' dictionary can be found across the world wide web, including even a version for PDAs or mobile devices; most of them are based on the Cologne version, which is freely downloadable for further academic use. Along with the Cologne searchable Tamil dictionaries, this project is one of the greatest contributions for the Indological community on the web. Another collection of digitalised dictionaries worth mentioning is the Digital Dictionaries of South Asia Project of the South Asia Language and Area Center at the University of Chicago,[2] which provides more than thirty high-quality online multilingual as well as monolingual dictionaries of the modern literary languages of South Asia plus the Anglo Indian "Hobson-Jobson".[3]

In a similar way, there are a growing number of Indology-related electronic libraries, i.e. libraries containing scanned printed materials. Unquestionably the most ambitious project in this field is the Indian Digital Library,[4] a joint effort of several dozen institutions coordinated by the Indian Institute of Science in Bangalore and the International Institute of Information Technology in Hyderabad. Its goal is to build a searchable online collection of one million books from public university libraries in India, predominantly in Indian languages, with a plan of supporting full text indexing and searching based on optical character recognition, or OCR.

One of the European counterparts to this effort is the GRETIL e-Library maintained by the State and University Library in Göttingen,[5] which is again a collection of electronic books on Indology-related subjects. Currently it contains more than 150 titles, mostly

1 http://www.sanskrit-lexicon.uni-koeln.de/

2 http://dsal.uchicago.edu/dictionaries/

3 Yule, Henry, Sir, Hobson-Jobson: *A glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive*. New ed., London 1903.

4 http://www.dli.ernet.in/

5 http://www.sub.uni-goettingen.de/ebene_1/fiindolo/gr_elib.htm

older standard works, and can be expected to grow substantially. Being a project of a German library, it offers a proper OPAC interface, which, we may hope, might become a standard for this kind of resource. The Göttingen State and University Library also maintains one of the largest repositories of searchable Sanskrit e-texts, the Göttingen Register of Electronic texts in Indian Languages, or GRETIL.[6] This is a collection of digitalised texts in Indian languages downloaded from reliable resources over the Internet, converted into a unified format and made available in three types of encoding, namely the UTF-8, or the Unicode, plus two older encodings, REE and CSX+. The whole collection contains nearly a thousand Sanskrit works, starting from the *Veda*, through the epics and selected *Purāṇas*, through *kāvya* to medieval texts, followed by e-texts in Pali and classical Tamil and an increasing number of texts in Hindi and Bengali.

The first machine-readable Sanskrit text, and perhaps one of the greatest contributions in this category of online Indology resources, was the digitalised *Mahābhārata* presented in the mid-1990s by Prof. Muneo Tokunaga from the Department of Indian Philosophy of the University of Kyoto, which was soon followed by the digitalised *Rāmāyaṇa*.[7] Both digitalised texts, based on critical editions,[8] were further proofread and modified by Prof. John Smith of Cambridge University in cooperation with the Bhandarkar Oriental Research Institute in Pune.

The development of Internet applications in general brought about some new trends in our area of interest as well, especially trends towards more interactive tools, such as searchable online databases of Sanskrit texts. At this point, I will just briefly introduce the database of Sanskrit e-texts and the database of Indian plant names developed as a part of the Pandanus project at the Seminar of Indian Studies of the Institute of South and Central Asia of the Faculty of Arts, Charles University in Prague,[9] under the guidance of Prof. Vacek. The original versions of the two partly interconnected databases were created during the initial stages of the Pandanus project in the late 1990s as a simple but useful tool for the study of the semiotics of nature in Indian literature. Since then, both databases have grown substantially. The Pandanus database of Sanskrit e-texts includes a collection of more than forty Sanskrit *kāvya* and *subhāṣita* texts with an online search interface. All of the texts were transcribed and proofread at the Institute by advanced students of Sanskrit. The Pandanus database of Indian plant names currently contains about 400 species and more than four thousand names in Latin, Sanskrit, Prakrit, Hindi, Bengali, Tamil, Malayalam and English. The details given for each species include the Latin name according to the current taxonomy along with its possible older Latin synonyms, the corresponding plant names in the above mentioned Indian languages and in English, a brief botanical description along with the plant's properties that may be important for its symbolic functions, a link to a list of occurrences of the plant's Sanskrit names in the Sanskrit e-texts database, an external link to a plant

---

6 http://www.sub.uni-goettingen.de/ebene_1/fiindolo/gretil.htm

7 Both texts are available online at http://www.cc.kyoto-su.ac.jp/~yanom/sanskrit/.

8 The Pune and the Baroda critical edition, respectively.

9 http://iu.ff.cuni.cz/pandanus/

image search service, selected data from relevant encyclopaedias and dictionaries, as well as other additional information, comments and notes.

However, there are a number of other approaches to the functionalities of Sanskrit text online applications. An excellent example of a corpus linguistics oriented approach is the Digital Corpus of Sanskrit, or DCS, a searchable collection of Sanskrit texts developed and maintained by Prof. Oliver Hellwig from the Department of Classical Indology at the University of Heidelberg.[10] It offers the possibility of searching for lexical units in a corpus of almost 200 Sanskrit texts containing more than two and half million manually tagged words. A set of references and a statistical evaluation of the occurrences in the texts can be generated for each lexical unit in the corpus, which allows the user, for example, to see the frequency of the appearance of a given word in time.

Another possible approach can be illustrated by a project called the "Gita Supersite" developed at the Department of Computer Science and Engineering of the Indian Institute of Technology, Kanpur,[11] by a team under the guidance of Prof. Prabhakar and Prof. Karnick. This is an example of a single text oriented website allowing a simultaneous display of the *mūla* text of a *Bhagavadgītā* verse along with one or more of the 13 Sanskrit commentaries, 2 Hindi commentaries and one English commentary that have been processed so far, not to mention one or more of 7 English translations and 3 Hindi translations, in one of nine Indian scripts or in Roman transliteration, which may be furthermore accompanied by the optional audio recording of the verse. The same model can be successfully applied to virtually any work of the Sanskrit commentarial traditions.

Yet another great example of an application intended as a tool for either guided or independent study of Sanskrit is the *Kramapāṭha* Reader presented by Dr. Peter Scharf from the Department of Classics at Brown University as a part of his larger Sanskrit Library Project.[12] The interactive Java application, which contains three Sanskrit texts, namely *Pañcatantra*, *Rāmopākhyāna* and *Aṣṭādhyāyī,* allows the student to choose how much of the very detailed and precise analysis of the Sanskrit text will be displayed for each verse, again accompanied by an optional audio recording. This approach can save the student considerable time and effort in learning the text, while at the same time it serves as a model of a proper text analysis.

The other sections of Dr. Scharf's Sanskrit Library Project already fall within the last of the three categories of Sanskrit online resources we want to cover in this article, namely grammar analysis tools, offering some very striking achievements in the field of Sanskrit computational linguistics, including language processing tools such as *sandhi* generators and *sandhi* analysers, morphological generators and analysers, Sanskrit corpora supporting advanced search options and Sanskrit parsers analysing the entered Sanskrit text into Sanskrit tagged hypertext with its morphology analysed, all of this available online without the necessity of installing

10 http://kjc-fs-cluster.kjc.uni-heidelberg.de/dcs/
11 http://www.gitasupersite.iitk.ac.in/
12 http://www.sanskritlibrary.org/

any further linguistic software etc. Let us mention three examples that offer more or less similar features with just slight variations; these include the website of the Computational Linguistic Research and Development Department at the Special Centre for Sanskrit Studies of the Jawaharlal Nehru University,[13] developed under the supervision of Dr. Girish Nath Jha, the website of the Department of Sanskrit Studies at the University of Hyderabad,[14] headed by Prof. Amba Kulkarni, and the Sanskrit Heritage Site of Prof. Gerard Huet from the French National Institute for Research in Computer Science and Control (INRIA).[15]

Without any overestimation of the importance of the Internet for Indian studies, let us conclude with a few reflections. There is no doubt that Indology has taken its place online and that it has already undertaken a considerable journey in utilizing the available technologies. The development of the form of its presence online may also be seen in the context of the ongoing discussion on Indological identities as one of the possible solutions to the issue of the perhaps desirable, yet disputable, phenomenon of a decreasing stress on a thorough knowledge of the Sanskrit language within the curriculum.

I dare say that within a few years – provided that current trends and efforts become coordinated – students of Indology with just intermediate command of Sanskrit will have at their disposal all the tools and resources necessary for an effective study of most of the important Sanskrit works online, i.e. a reliable edition of the text, a reliable dictionary and a detailed grammar analysis tool for use as and when needed. Of course, a proper graduate student of so-called classical Indology, if the subject survives as such, should not and will not depend on any of these tools.

Interestingly enough, this very intersection of information technologies and Sanskrit linguistics may at the same time be one of the points where, if I am once again allowed the terms, Western and Indian Indology can meet, and where they actually have already been meeting for some years.[16] We may see a growing interest in computational linguistics in India with maybe even some practical overlaps for European students. Unfortunately, experience suggests that this combination of interest in corpus linguistics, computational linguistics and Sanskrit linguistics is, in our longitudes, a rather rare combination for students of Indology, some of whom have quite different levels of problems, being sometimes unable to use, after a few years of study, a proper font to display the *devanāgarī* script or even the diacritics in Roman transliteration. The above mentioned interest could, or perhaps even should, be encouraged by introducing the available resources to Sanskrit students during the initial stages of their study, or even making them use such resources actively. However, I would like to leave the question of the desirable computer and internet skills of Indology students, particularly in our part of Europe, open for further consideration.

---

[13] http://sanskrit.jnu.ac.in/

[14] http://sanskrit.uohyd.ernet.in/

[15] http://sanskrit.inria.fr/

[16] Annual International Sanskrit Computational Linguistics Symposia have been held since 2007 (Paris, Providence, Hyderabad, New Delhi).

E-REFERENCES

http://www.sanskrit-lexicon.uni-koeln.de/
http://dsal.uchicago.edu/dictionaries/
http://www.dli.ernet.in/
http://www.sub.uni-goettingen.de/ebene_1/fiindolo/gr_elib.htm
http://www.sub.uni-goettingen.de/ebene_1/fiindolo/gretil.htm
http://www.cc.kyoto-su.ac.jp/~yanom/sanskrit/
http://iu.ff.cuni.cz/pandanus/
http://kjc-fs-cluster.kjc.uni-heidelberg.de/dcs/
http://www.gitasupersite.iitk.ac.in/
http://www.sanskritlibrary.org/
http://sanskrit.jnu.ac.in/
http://sanskrit.uohyd.ernet.in/
http://sanskrit.inria.fr/