The construction of IndoWordNet will, no doubt, open up a new vista of lexical resources for most of the Indian languages. It will eventually become an open resource usable in all works of descriptive, applied, and computational linguistics.[6]

8.3.1.2.2. Sanskrit
By Amba Kulkarni

8.3.1.2.2.1. Electronic corpora

Classical languages like Sanskrit in addition to being a repository of classical literature play an important role as carriers of culture and history. With the advent of the electronic age, it was natural that enthusiasts as well as academics took interest in preserving the texts using new technology. Like other Indian languages, Sanskrit also experienced its initial hurdles as regards keyboard, fonts, etc. Scharf & Hyman 2012 examines fundamental linguistic issues in encoding Sanskrit texts and discusses processing principles relevant to the present technology.

The birth of the Internet paved the way for enthusiasts to make their individual collections public. Thanks to the World Wide Web technology, numerous digital collections of Sanskrit could be made available online. The critical edition of the Sanskrit Epic Mahābhārata, based on John Smith's revision of Muneo Tokunaga's version of the text, is made available online by the Bhandarkar Oriental Research Institute (http://bombay.indology.info). George Cardona headed a project to create a databank and relational database of major Sanskrit grammatical texts such as the Aṣṭādhyāyī, Mahābhāṣya, Kāśika, Prātiśākhyas, and Nirukta (http://sanskrit-library.org). Another major contribution is the bibliographical listing of the philosophical literature of India during its classical phase and also the secondary material in European languages on this literature at http://faculty.washington.edu/kpotter/ckeyt/home.htm, Karl Potter being the General Editor.

The year 1999–2000 (Kaliyugabda 5101) was declared as the "Year of Sanskrit" by the Government of India, and during this year, a major initiative, SanskNet, was taken up by the Rashtriya Sanskrit Vidyapeetha, Tirupati, under the leadership of K. V. Ramakrishnamacharyulu. SanskNet brought together various institutes of traditional learning, oriental research institutes, manuscript collection centers, and libraries to develop a digital corpus of Sanskrit, and make it available online to researchers. In the first phase, around 475 books were digitized. The mammoth task of digitizing Sanskrit texts is being continued further by the Rashtriya Sanskrit Sansthan.

---

[6] Editorial note: Chaplot, Bhingardive & Bhattacharya 2014 presents a graphical user interface to browse and explore the IndoWordNet lexical database for various Indian languages.

The Göttingen Register of Electronic Texts in Indian Languages (GRETIL), a comprehensive repository of e-texts in Sanskrit and other Indian languages, was started by Reinhold Grünendahl, with the intention of its being a 'cumulative register of the numerous download sites for electronic texts in Indian languages' (http://gretil.sub.uni-goettingen.de/). Rather than scanned books or typeset PDF files, these documents are plain texts, in a variety of encodings, and are machine-readable, so that (for instance) word search can be performed on them.

Searches become even more effective if the text is semantically marked. SARIT (http://sarit.indology.info/), an initiative by Dominik Wujastyk, hosts texts that are marked up (tagged) using the rich Text Encoding Initiative (TEI) system and provides a Search and Retrieval facility for them.

8.3.1.2.2.2. Lexicon development

In the late 1990s the Cologne Digital Sanskrit Lexicon (CDSL) project (http://www.sanskrit-lexicon.uni-koeln.de/) undertook the digitisation of major bilingual Sanskrit dictionaries of the 19th century. Major works such as Monier-Williams' *Sanskrit-English Dictionary*, Böhtlingk and Roth's *Sanskrit-German Dictionary*, Apte's *English-Sanskrit Dictionary*, Wilson's *Sanskrit-English Dictionary*, and Stchoupak's *Sanskrit-French Dictionary* have been digitised and are available online with various search features. The search features were developed in collaboration with Peter Scharf's Sanskrit Library Project (http://sanskritlibrary.org). There was a similar effort at Rashtriya Sanskrit Vidyapeetha Tirupati by Varakhedi and his team in the last decade to digitise the Sanskrit encyclopedic lexicon *Vācaspatyam* (Varakhedi & Jaddipal 2009). While the knowledge structure of Amarakośa has been explored by computer scientists (http://sanskrit.uohyd.ernet.in/scl/amarakosha/), Sanskrit scholars have also taken the initiative in the development of modern e-lexicons such as WordNet (http://www.cfilt.iitb.ac.in/wordnet/webswn/) under the guidance of Malhar Kulkarni.

Sanskrit being highly inflectional and productive in its derivational morphology, accessing dictionaries such as Monier-Williams is laborious and needs a good knowledge of grammar. The free word order and the ambiguities at various levels of analysis make it quite difficult to understand a Sanskrit text. Current technology provides a mechanism through which one can bring some relief to the life of a Sanskrit reader. The late 1990s and early years of the 21st century saw various individual efforts towards the development of Sanskrit computational tools. The first parser for Sanskrit using Integer Programming was developed by Pushpak Bhattacharyya in 1987 as a part of his M.Tech. dissertation at IIT-Kanpur. In the early 1990s, the Center for Development of Advanced Computing (CDAC) and the Academy for Sanskrit Research, Melkote, developed morphological generators. An ambitious research and development activity for Sanskrit and other Indian

languages was started by Girish Nath Jha at Jawaharlal Nehru University, Delhi (http://sanskrit.jnu.ac.in). Amba Kulkarni from the Akshar Bharati group, who was actively engaged in developing language accessors among Indian languages taking insights from Pāṇini's Grammar, joined hands with K. V. Ramakrishnamacharyulu and Srinivas Varakhedi of Rashtriya Sanskrit Vidyapeetha and started developing an accessor for Sanskrit.

Gérard Huet, while developing a hypertext Sanskrit-French dictionary, strongly felt the need of linking the words to the morphological generator; this resulted in the implementation of a Sanskrit computational linguistics platform conceived as a coordinated set of Web services around a lexical database obtained mechanically from his highly structured Sanskrit Heritage dictionary (http://sanskrit.inria.fr). A typical Sanskrit reader needs access to a dictionary, as well as grammar. The Sanskrit Heritage site provides a user with a hypertext dictionary where the words are linked to the morphological generator, a segmenter to split a continuous Sanskrit text into words, and a word-level and a sentence-level analyser under one integrated environment.

Peter Scharf (http://sanskritlibrary.org) developed a 'digital library dedicated to facilitating education and research in Sanskrit by providing access to digitized primary texts in Sanskrit and computerized research and study tools to analyze and maximize the utility of digitized Sanskrit text.' The Sanskrit Library is also engaged in aligning digital texts with digital images of manuscripts which allows immediate focused access to particular passages in manuscripts and conversely, when examining a manuscript, to complementary texts, metadata, linguistic tools, and lexical resources.

The Digital Corpus of Sanskrit (DCS) (http://kjc-fs-cluster.kjc.uni-heidelberg.de/dcs/index.php), developed by Oliver Hellwig, is another major collection of Sanskrit texts; it provides access to a searchable collection of lemmatized Sanskrit texts and to the database of the SanskritTagger software. With the SanskritTagger it is possible to analyze unprocessed digital Sanskrit text both lexically and morphologically. The DCS has been designed to support research in Sanskrit philology. It provides search of lexical units and their collocations from a corpus of around 3,000,000 words.

Collaboration between the Institute for Research in Computer Science and Automation (INRIA) in Paris and the Department of Sanskrit Studies of the University of Hyderabad provided a common platform to researchers working in the field of Sanskrit computational linguistics in the form of a symposium. In October 2007, the joint team organized the First International Sanskrit Computational Linguistics Symposium at INRIA. This was followed by a series of symposia — at Brown University (May 2008), University of Hyderabad (Jan. 2009), JNU, Delhi (Dec. 2010), and IIT Bombay in Jan. 2013. The symposium publications have been published as Huet, Kulkarni & Scharf (eds.) 2009, A. Kulkarni & Huet (eds.) 2009, Jha (ed.) 2010, and M. Kulkarni & Dangarikar (eds.) 2013.

The Sanskrit Consortium of seven institutes[7] in India developed tools for analysis of Sanskrit texts, viz. Samsaadhanii (http://sanskrit.uohyd.ernet.in/scl); http://tdil-dc.in/san/). The tools contain a morphological analyser and generator, tools for segmentation and formation of sandhi, compound processors, a parser leading to kāraka analysis and a machine translation system from Sanskrit into Hindi.

Gaveśikā, a search engine for Sanskrit, was launched by the University of Hyderabad in March 2012. This search engine integrates a Sanskrit morphological analyser with the basic search engine resulting in better search results.

Efforts are also in progress to integrate Peter Scharf's Sanskrit Library, Gérard Huet's Heritage segmenter, and Amba Kulkarni's parser (P. Goyal, Huet et al. 2012). In addition to developing various tools, the consortium under the guidance of K. V. Ramakrishnamacharyulu has developed tagsets and guidelines for annotating Sanskrit compounds and developing Sanskrit treebanks following Pāṇinian grammar. A treebank of around 4,000 sentences following these guidelines has been developed.

Developing computational tools for Sanskrit poses many challenges. Though Sanskrit has a rich grammatical tradition, this rich literature is not readily digestible by a computer scientist working on Sanskrit. It is only the efforts of Cardona (1988), Gillon (1995), Kiparsky (2009), Staal (1967), and others throwing light on various aspects of Pāṇinian grammar such as the organisation and interpretation of Pāṇini's sūtras, presenting Pāṇini's formal grammar in modern linguistic terminology, and describing the structure of the Sanskrit language, that came to the rescue of computational linguists engaged in bridging the gap between modern technology and the enormous descriptive grammatical literature of the Pāṇinian tradition.

Pāṇini's Aṣṭādhyāyī is often compared to a computer program, for its rigour and coverage. No wonder there are notable efforts in "modeling" Pāṇini by Swami Sri Taralabalu (http://www.taralabalu.org/panini/), Mishra (2009), Scharf (2009), P. Goyal, Kulkarni & Behera (2009), and Subbana & Varakhedi (2010).

[7] International Institute of Information Technology-Hyderabad, Jagadguru Ramanandacharya Rajasthan Sanskrit University-Jaipur, Jawaharlal Nehru University-Delhi, Poornaprajna Vidyaapetha-Bangalore, Rashtriya Sanskrit Vidyapeetha-Tirupati, Sanskrit Academy Osmania University-Hyderabad, and University of Hyderabad-Hyderabad.