

# संस्कृत संज्ञानिकी

## Knowledge Technology for Sanskrit

June 2019

[KTforSanskrit@bvbdelhi.org](mailto:KTforSanskrit@bvbdelhi.org)

*Patronage:* Ashok Pradhan

*Editing:* Shashiprabha Kumar & Om Vikas

*Coordination:* Pushpendra Sharma

**It is proposed to organize a workshop on *Knowledge Technology for Sanskrit: R&D prospects on 11<sup>th</sup> October 2019 at Bharatiya Vidya Bhavan, Delhi Kendra.*** Bharatiya Vidya Bhavan was founded in 1938 to promote *Sanskriti* (Indian Culture) And *Sanskrit*, the mother of languages.

We welcome short notes of about 200 words describing achievements and prospective research directions in the field of *Knowledge Technology for Sanskrit*.

न हि ज्ञानेन सदृशं पवित्रमिह विद्यते (गीता ४.३८)

“Knowledge is experience: Everything else is just information.” . . . Albert Einstein

Knowledge is residue of thinking. It is created at the boundary of old. Knowledge flows from community to community spatially and temporally. Knowledge links to network of ideas, memory, predictions, procedures, beliefs, cultural expressions and experiences. Sanskrit scriptures have rich treasure of linguistic, philosophical, cultural, scientific and technological knowledge.

Language technologies and tools have been developed under the TDIL program of Min. of Electronics & IT, Government of India to process information in Indian Languages, and *a few technologies for Sanskrit too.*

*There are still technological issues* relating to input, parsing, word processing, annotated corpora, dictionaries, knowledge representation, encoding with Vedic symbols, digitization of manuscripts, Sanskrit WordNet, translation memories, expert systems, Authoring system, spelling check, grammar check, searchable Sanskrit dictionaries, content creation, e-learning resources, etc. Requisite tools need to be developed.

There is scope of further *research in the field of speech processing* using phonetic rules as prescribed in Shiksha Shastra, metre study as in Chhand Shastra.

A word in Sanskrit may connote wider context that interlinks other contextual concepts for example lotus, water. It is possible to develop Sanskrit based *Concept Networking Language (CNL)* that may become Universal Networking Language (UNL). CNL will be useful in developing translation systems among world languages.

Semantic web, cognitive computing, brain simulation, building consciousness are *emerging research areas* wherein Sanskrit studies may help.

There is need to identify tools and resources for *capacity building* at Sanskrit schools, Pathshalas, colleges, universities and centers of excellence for Sanskrit studies. There is need to develop e-learning resources in open domain (MOOCs), language laboratory software, searchable linguistic resources, etc.

*Standards* may be evolved for input mechanism, script-grammar, language grammar, transcription tables, phonetic table etc. *Standard for Devanagari based Indian Phonetic Alphabet (IPA)* may be developed that will greatly be useful in speech processing and Linguistic studies of especially Indian Languages and some world languages.

*Synergistic collaboration between Government, Academia and Industry* would be required to accomplish the desired objectives. It is desirable to promote Sanskrit based studies worldwide. All the technologies and tools developed through Governmental support and initiatives need be available in open source for use by all.

Broad areas for discussion may include

- (i) **Input-Output** (keyboard, voice, OCR, thought driven, display (print, voice, sign) for normal and impaired users)
- (ii) **Processing tools** (parser, spell checker, grammar checker, summarizer, etc.)
- (iii) **Linguistic data resources** (Lexicon, Linguistic Data resource, Ontologies, etc.)
- (iv) **Standards** (IPA, CNL, etc.)
- (v) **Applications** (Education, Health care, Raga based therapy, heritage, etc.)
- (vi) **Capacity Building** (traditional technology-enabled Learning practices, pedagogy, teachers training, Massive Open Online Courses, etc.)
- (vii) **Sanskrit based Cognitive Computing** (Building Consciousness, Brain Simulation, Augmented Intelligence, Cognitive Augmentation, etc.)
- (viii) **Strategy for Synergistic Collaboration between Government, Academia and Industry.**

We solicit your comments, suggestions and possible futuristic projects on the above subject.

## Sanskrit Resources and Tools Available on TDIL-DC Portal

Apart from Sanskrit software tool CDs at [www.ildc.in](http://www.ildc.in) various other Sanskrit tools and resources are developed under TDIL programme are hosted and made available for free download for research community at [www.tdil-dc.in](http://www.tdil-dc.in) are listed below with brief description.

### Online Sanskrit Tools:

- 1. Transliteration Tool:** Sanskrit is written in several scripts and using various romanized transliterations. This tool provides conversion facility among the Romanized transliteration schemes into Devanagari and also from Devanagari into various romanised transliteration schemes.
- 2. Morphological Generator:** The Morphological Generator shows the subantas and tingantas (inflectional forms of a noun or a verb). It also handles derivational morphology showing kridanta and taddhita forms.
- 3. Morphological Analyzer:** Analyses a Sanskrit word giving its nominal stem / verbal stem along with its various linguistic features such as Lexical-category, Gender, Number, Case, Person, etc.
- 4. Sandhi Splitter:** This shows various possible splittings (pada-patha) of a given Sanskrit string, and also the various components in case of a compound. More than one ways of splits are ranked by machine based on the frequency of occurrence in the manually sandhi-split corpus.
- 5. Sandhi:** This tool joins two Sanskrit words following the Paninian sutras.
- 6. Sanskrit E-learning and Multimedia:** E-learning portal for Sanskrit developed by JNU.

### Resources:

- 1. Sanskrit SakalBharati Unicode Font**
- 2. Bilingual Sanskrit–Hindi Apte Dictionary** (This Apte Dictionary is tagged and bilingual (Sanskrit, Hindi) with 100,000 words.)
- 3. Sanskrit Monolingual Text Corpora**
- 4. Sanskrit -Amarakosha Head Words** (Amarakosha is the oldest Sanskrit thesaurus authored by Amarasimha giving the synonymous words for Sanskrit words. Total head words are 11582 in plain text format.
- 5. Sanskrit- synset (WordNet):** Developed, for each synset a POS category, example and set of synonyms are defined. It is a synonym set and provided in \*.syns file formats. Approx. 23140 synsets have been developed

Further details please contact [VKumar@meity.gov.in](mailto:VKumar@meity.gov.in)