

Introduction to Sanskrit Shabdmitra: An Educational Application of Sanskrit Wordnet

Malhar Kulkarni, Nilesh Joshi, Sayali Khare, Hanumant Redkar, Pushpak Bhattacharyya
Center for Indian Language Technology
Indian Institute of Technology Bombay
malharku@gmail.com, joshinilesh60@gmail.com, sayali.khare92@gmail.com,
hanumantredkar@gmail.com, pushpakbh@gmail.com

Abstract

This paper introduces a digital tool, viz., Sanskrit Shabdmitra, for learning and teaching of Sanskrit in active classroom environment as well as in other formal and informal set-up. It is based on an existing digital resource called Sanskrit Wordnet created at IIT Bombay. Thus, this paper also describes a direct application of Sanskrit Wordnet in particular, and Wordnet in general in the education domain. It describes the structure and various features of Sanskrit Shabdmitra.

1 Introduction

Sanskrit Wordnet¹ (SWN) was created at IIT Bombay as a major and unique lexical resource of Sanskrit (Kulkarni et al., 2010a). Kulkarni (2017) describes this effort in detail and demonstrates the contribution it made to the digital lexical resources of Indian languages. The effort of enriching SWN continues and scholars have tried to study it from the point of view of various natural language processing (NLP) tasks. Bhingardive et. al. (2014) developed well researched method based on SWN to populate one Wordnet using another lexical resource and Redkar et. al. (2016) have developed tool to populate one synset using two synsets with the help of SWN. Wordnet² and IndoWordNet³ has been used at various NLP tasks and applications. One such application is 'IndoWordNet::Similarity' developed by Bhingardive et. al. (2016) which measures semantic similarity and relatedness between two synsets in IndoWordNet. Similarly, IndoWordNet has been used for tasks such as Word Sense Disambiguation (Bhingardive and Bhattacharyya, 2017) for finding the most frequent sense using word and sense embeddings. This justifies the importance of IndoWordNet for word sense disambiguation for Indian languages. Similar to this, Sanskrit Wordnet can be used for the development of such tools, methods and utilities. Further, SWN can be helpful in explaining तत्सम (tatsama) and तद्भव (tadbhava) words which appear in any Indian languages. In this way Wordnet as a resource can be useful in many NLP tasks. Can Wordnet also be used as a base in creating an educational tool to teach and learn language? YES. We found that Wordnet can certainly be used as a base to create a tool to teach and learn Sanskrit. In this paper, in what follows, we elucidate how Sanskrit Wordnet can be used to develop educational application for teaching and learning Sanskrit language. Thus, a digital aid, Sanskrit Shabdmitra, has been introduced in this paper.

The paper is organized as follows - section 2 provides the literature survey; section 3 briefly mentions the related work; section 4 introduces the Sanskrit Shabdmitra, its structure, and its features in detail, explains how Shabdmitra enriches Wordnet, provides some applications; next section concludes the paper; this is followed by the future work.

¹<http://www.cfilt.iitb.ac.in/wordnet/webswn/wn.php>

²<https://wordnet.princeton.edu/>

³<http://www.cfilt.iitb.ac.in/indowordnet/>

2 Literature Survey

Sanskrit, belonging to the Indo-Aryan family of languages, is one of the ancient languages in the world. There is a rich tradition of developing a vast vocabulary in Sanskrit literature (Kulkarni et al., 2010a). Most of the languages in the Indo-European language family can be traced back to Sanskrit (Kulkarni et al., 2010b). There are various grammatical features and properties of Sanskrit which may not be present in other Indian languages (Redkar et al., 2014).

With the increase in the digital presence across the globe, content digitization and digital language learning have been growing enormously. Vocabulary is a crucial part of language learning. Learning Sanskrit vocabulary is one of the challenging tasks for any language learner. There are several applications and platforms available for curriculum based education, but very few are meant for language learning and active classroom. The Indian government is now supporting digital education and has taken several steps in digital language education. Following are government-driven platforms in digital language education:

- NCERT⁴ provides e-textbooks and supplementary books for students. It also provides guidelines for teachers for effective teaching.
- NROER⁵ is a Pan-Indian collaborative platform for teachers, students and professionals from various educational institutes. It allows uploading the digital content such as articles, text, poems, etc. which can be publicly available to the internet users.
- Swayam⁶ is another government designed program, collaborating with several government organisations, such as UGC⁷, AICTE⁸, NCERT, IGNOU⁹, etc. It covers courses from secondary education to post graduation. It teaches subjects like English, Hindi, and Sanskrit through video lectures and provides reading material, self-assessment tests, etc., and has an online discussion forum.

Apart from the above, there are some other non-government platforms engaged in digital language education. They are as follows:

- Openpathshala¹⁰ is an online platform for Sanskrit language teaching using lessons and video tutorials for learning Sanskrit grammar.
- pANini aShTaadhyaayii sUtra paaThaH¹¹ contains the audio pronunciation of the entire treatise on Sanskrit grammar (8 chapters of sūtras), called aṣṭādhyāyī by maharṣi pāṇini.
- shaale¹² provides the traditional methods of teaching Sanskrit using videos, live streaming (webcast), video on demand, audio documentation service, etc.
- Sanskrit Documents¹³ has the vast variety of documents which provides a collection of various links to various repositories useful for Sanskrit language learning.
- Vyoma¹⁴ introduces a guide of Sanskrit to generate a sentence, viz., Sanskrit vocabulary builder, Sanskrit pronunciation, Yogasutraparichaya, Saptāhastotra Saṅgrahaḥ, Sanskrit games, Learn Sanskrit through Hindi and English, etc.

⁴<http://ncert.nic.in/>

⁵<https://nroer.gov.in/>

⁶<https://swayam.gov.in/>

⁷<https://www.ugc.ac.in/>

⁸<https://www.aicte-india.org/>

⁹<http://www.ignou.ac.in/>

¹⁰<https://openpathshala.com/>

¹¹<http://surasa.net/music/samskrta-vani/ashtadhyayi.php>

¹²<https://www.shaale.com/>

¹³http://sanskritdocuments.org/learning_tools/index.php

¹⁴<http://www.sanskritfromhome.in/>

- learnsanskrit.org¹⁵ aims to teach Sanskrit grammar, providing a generative grammar guide of Sanskrit.
- Push to learn¹⁶ is a platform where students learn vocabulary from the school's course-books. However, this platform is not meant for Sanskrit.
- Spoken tutorial¹⁷ offers self-paced, multi-lingual courses. Anybody with a computer and a desire for learning can access this platform.
- Robomate¹⁸ is a curriculum based language learning app which has interactive study material for students like attractive video lessons.
- Byju's¹⁹ is a platform for interactive learning consisting of video lessons for Science, Maths, Economics and Business studies for school education. However, this platform does not have language learning facility.
- Duolingo for Schools²⁰ is a blended learning mate for the classrooms. Duolingo lessons provide personalized feedback to each student and help them to get the most out of classroom instruction. It also provides language specific class tips for teachers; such as phonetic inventory of a language, morphology, syntactic and semantic information. However, this tool does not facilitate Sanskrit language learning.

Other online resources for Sanskrit are bilingual dictionaries and thesauri which provide only the meanings of the words, such as Monier-Williams Dictionary²¹, Apte's Dictionary²², Spoken Sanskrit Dictionary²³, etc. Apart from these, there are some online dictionaries and thesauri in Sanskrit viz., Amarakosha²⁴, Sabda-kalpadruma²⁵, Vacaspatyam²⁶, etc. These online resources have domain-specific ontology, i.e., mythological ontology. Whereas, Wordnet does has been considered an upper ontology (Navigli and Velardi, 2004).

Most of these tools and platforms are in the form of text material, presentations, videos, lesson plan, etc. However, they do not provide relational semantics. Majority of them are not interactive and curriculum specific vocabulary learning is not available. It should be noted that one common thing among all the above resources is that they are more focused on individual learning and do not provide the active classroom learning. This is the desideratum as the knowledge of words or concepts in Sanskrit is not available as per the school curriculum. On the other hand, Sanskrit Shabdmitra, introduced here, is a digital language learning platform designed for Sanskrit vocabulary learning as per the school curriculum and for individual learning as well. This shall be explained in detail in section 4.

3 Related Work

Semantic relations of words helps in better understanding of new vocabulary (Lin, 1997). One such rich lexical resource based on semantic relations is viz., the Princeton WordNet²⁷, i.e., the WordNet(Miller, 1995), has been explored for vocabulary learning and other language learning applications (Hu et al., 1998; Sun et al., 2011; Brumbaugh, 2015; Hiray, 2015). Recently,

¹⁵<http://learnsanskrit.org>

¹⁶<http://pushtolearn.com/features>

¹⁷<https://spoken-tutorial.org>

¹⁸<https://roboestore.com/>

¹⁹<https://byjus.com/>

²⁰<https://schools.duolingo.com>

²¹<http://www.sanskrit-lexicon.uni-koeln.de/monier/>

²²<http://www.aa.tufs.ac.jp/~tjun/sktdic/>

²³<http://spokensanskrit.org/>

²⁴<https://sanskritdocuments.org/sanskrit/amarakosha/>

²⁵<http://www.sanskrit-lexicon.uni-koeln.de/scans/SKDSscan/2013/web/webtc2/index.php>

²⁶<https://archive.org/details/vacaspatyam02tarkuoft>

²⁷<https://wordnet.princeton.edu/>

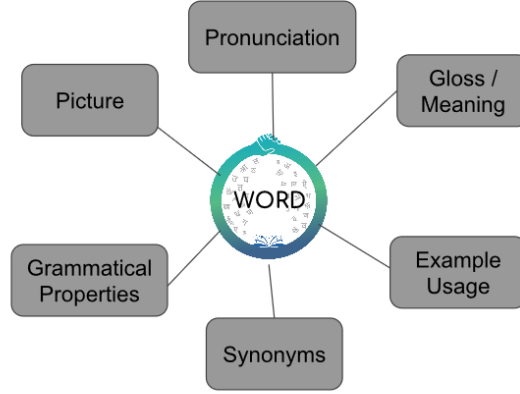


Figure 1: Shabdmitra as a friend of a word by providing word-meaning, example usage, pronunciation, picture, synonyms, other grammatical features, etc.

Hindi Wordnet²⁸ (HWN) has been used to build a teaching and learning digital aid, Hindi Shabdmitra, for Hindi language education in formal (schools) and informal (self-learning) setups (Redkar et al., 2017a). Additionally, the development of Marathi Shabdmitra, using Marathi Wordnet as a resource, is also under process.

A study of current digital resources used by the various educational institutions was also done as part of the background study. The outcome showed that there is a lack of quality resources which can cover all aspects of language learning such as grammar, concepts, usage, and pronunciations in an effective manner.

This motivated us to develop a digital aid, viz., Sanskrit Shabdmitra, that would fill this gap for Sanskrit language teaching and learning in both formal and informal learning environment.

4 Sanskrit Shabdmitra: an educational application using Sanskrit Wordnet

4.1 Shabdmitra

Shabdmitra is an umbrella of multilingual digital aid of language teaching and learning for Indian languages. It is built using IndoWordNet (Bhattacharyya, 2010) as a resource and is related to Hindi Shabdmitra (Redkar et al., 2017b), which is an initiative of IIT Bombay, India²⁹, exploring the applications of wordnet in education domain. The term Shabdmitra and its meanings were originally conceived by Malhar Kulkarni. The term Shabdmitra, शब्दमित्र is coined from two words 'shabda', शब्द, i.e., 'a word' and 'mitra', मित्र, i.e., 'a friend'; also means 'the Sun'. Therefore, Shabdmitra means a friend which helps in understanding a given word/concept. Using the second meaning of the word 'mitra' mentioned above, the word Shabdmitra would mean an illuminator of a word or concept. Thus, the function this tool aims to perform and the goal it wants to achieve is aptly expressed by the word 'Shabdmitra' itself. Thus, this term 'Shabdmitra' can be called self-explanatory. (anvartha-samjñā). This has been visualised in figures 1 and 2.

In Shabdmitra, the IndoWordNet data such as gloss, example sentence(s), synonyms and lexico-semantic relations are used and further augmented in order to cater to language learning needs. It is proposed to develop Shabdmitra for 18 Indian languages viz., Assamese, Bodo, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Manipuri, Malayalam, Marathi, Nepali, Odia, Punjabi, Tamil, Telugu, Urdu and Sanskrit, which are present in the IndoWordNet³⁰.

²⁸<http://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php>

²⁹<http://www.iitb.ac.in/>

³⁰<http://www.cfilt.iitb.ac.in/indowordnet/>



Figure 2: Shabdmitra as an illuminator for a word where it provides multiple senses, lexico-semantic and ontological relations, etc. of the same word

Figure 3 illustrates the IndoShabdmitra for IndoWordNet languages.

Shabdmitra is a multifaceted model which acts as a platform, as a resource and as a brand for the multilingual Indian scenario.

- As a Platform, various Indian languages which are present in IndoWordNet are made available at a single place.
- As a Resource, the multilingual Shabdmitra can be easily developed using the shared and not-shared data available in all the wordnets in the IndoWordNet database.
- As a Brand, all the wordnets can be branded under the umbrella of Shabdmitra which can be seen in figure 3.

Synset Category
Common
Uncommon
Common in Indian languages
Region and Language Specific

Table 1: Classification of Synsets by Bhattacharyya (2010)

4.2 Sanskrit Wordnet

Wordnet is a lexical resource composed of synsets, lexico-semantic relations and ontological information. Synset is the basic building block of a wordnet and it contains a gloss, an example sentence and synonyms. Wordnet is linked by semantic relations like hypernymy-hyponymy (is-a), meronymy-holonymy (part-of), troponymy (manner-of), etc. and by lexical relations like antonymy, gradation, etc. (Bhattacharyya, 2017). IndoWordNet is a linked structure of wordnets of 18 Indian languages from Indo-Aryan, Dravidian and Sino-Tibetan language families (Bhattacharyya, 2010).

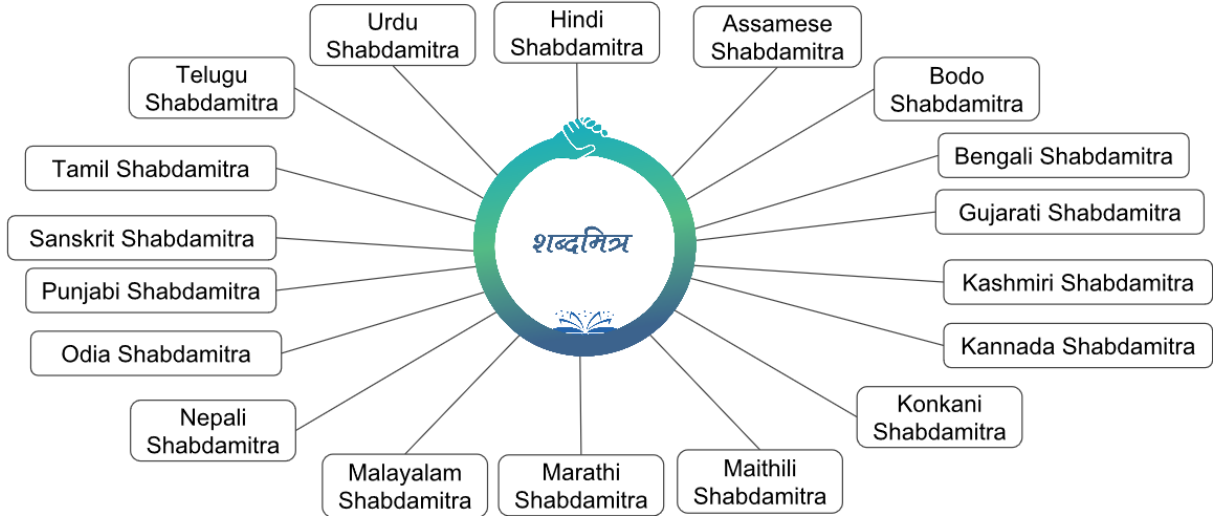


Figure 3: IndoShabdmitra: Shabdmitra of all IndoWordNet Languages

Sanskrit Wordnet is a part of IndoWordNet and is constructed using an expansion approach in which Hindi wordnet is used as a source (Kulkarni et al., 2010a). As Sanskrit has both the vedic as well as the modern literature, it has a greater scope of vocabulary than that of Hindi. Therefore, though the Sanskrit wordnet was built using an expansion approach from HWN, all Sanskrit synsets could not be developed. Hence, Sanskrit was developed in versions. And this development is an ongoing process. In this approach, the following part-of-speech wise method has been adapted for creating synsets (Kulkarni et al., 2010a):

Nouns

In the case of nouns in Sanskrit, a gender information is included in the word itself. In SWN, all nouns are stored in nominative singular form, however other Indian WordNets store nouns in their root forms. For example, देवः (devah), मतिः (matih), etc. are stored in nominative singular form in SWN; while in HWN, देव (deva), मति (mati), etc. will be stored in root form.

Adjectives

Adjectives in general have no gender of their own. However, in Sanskrit, they take the gender of the nouns which they qualify. Hence, in the synsets of adjectives, only the root forms are included. For example, भद्र (bhadra), निर्मल (nirmala) are stored in root form.

Adverbs

In SWN, adverbs are in their root form, however, it is observed that some of the adverbs have case-ending suffixes. These suffixes indicate the closed form of the word in that particular case-ending. Hence, these adverbs are regarded as frozen adverbs. In such cases, adverbs are stored with their case-ending suffixes. For example, व्यतिरेकेण (vyatirekeṇa) is stored as instrumental singular form; रहसि (rahasi) is stored as locative singular form. While, साधु (sādhu) is stored in root form.

Verbs

In SWN, verbs are also stored in their root form. For example, भू (bhū); कृ (kṛ); are stored in root form.

Apart from the above parts-of-speech information, other information such as gloss, examples are stored in the SWN database. Synsets in wordnet are interlinked by means of conceptual semantic and lexical relations. A combination of the glosses given in traditional dictionaries like Shabdakalpadruma³¹, vācyaspatyam³² and the translation of the gloss of the HWN synset is used to create SWN gloss for nouns, adjectives and adverbs. In the case of verbs, though these traditional Sanskrit dictionaries contain etymology based glosses, they are not appropriate for verbs which has ontology based wordnet structure. Hence, Navyanyāya terminology has been adapted for verbal glosses to construct synsets (Kulkarni et al., 2010b).

All these data, features and properties of SWN can be effectively used and utilised for teaching and learning Sanskrit. This is the base for building Sanskrit Shabdmitra which uses SWN data for language education purpose. This has been explored in detail in the next section.

4.3 Sanskrit Shabdmitra: Structure, Features and Applications

Sanskrit Shabdmitra (संस्कृत शब्दमित्र) is a digital language teaching and learning tool for Sanskrit language education. It uses Sanskrit Wordnet (SWN) as a resource. SWN was originally developed for the research purpose in the area of natural language processing. Soon, it was realized that this rich resource can be applied and used in developing educational applications. Sanskrit Shabdmitra is one such application of SWN.

Shabdmitra has been devised by taking into consideration the various stakeholders of this application. The major stakeholders of Sanskrit Shabdmitra are: Teachers, Students and Parents. Teachers' concern is that he/she should be able to convey the entire content to students in all the possible nuances and make them competent in language learning, and prepare them for examinations. Students' concern is that he/she should learn and understand the content as exhaustively as possible in all nuances and grow in terms of competence and be prepared for examinations. Parents' concern is that their child should get quality education and obtain competitive results. All these stakeholders and their concerns were considered while designing this digital aid.

4.3.1 Structure

In this tool, SWN data, features and properties are further augmented, simplified and presented in the form of educational application in order to cater to language teaching and learning requirements of Sanskrit language education.

SWN data such as gloss, example(s), synonyms, ontological information, lexico-semantic relations, etc. forms the content of Sanskrit Shabdmitra. Some of this information is customized and modified as per the language learning requirement and the learning levels of the individuals. Apart from this, Sanskrit Shabdmitra has various other features which are stored in the Shabdmitra database. The details of these features are presented in the next section.

Broadly, Sanskrit Shabdmitra has two types of interfaces, viz., Class-Wise and Level-Wise. Following sections elaborate on the same:

Class-wise interface

Class-Wise interface is designed specifically for classroom or formal setup wherein Sanskrit teacher uses this digital aid. Here, the data is presented in the interface, lesson by lesson. In this interface, the teacher chooses a school curriculum board (CBSE, ICSE, State Board, etc.); followed by a class to which he/she wants to teach; followed by a lesson/chapter. Once he/she clicks on a chapter, all the words from that chapter appear in the order in which they appeared in the textbook. While teaching, teacher can simply click on any of the word from the list and the word-specific information with the same sense is displayed accordingly in the interface. In most schools in India, Sanskrit is considered as second or third language. Hence, students have Sanskrit as a subject in the secondary. Therefore, the provision is made to include Sanskrit as

³¹<https://www.sanskrit-lexicon.uni-koeln.de/scans/SKDSscan/2013/web/webtc2/index.php>

³²<https://www.sanskrit-lexicon.uni-koeln.de/scans/VCPScan/2013/web/webtc2/index.php>

a 2nd or 3rd language in the school setup. Figure 5 shows the class-wise interface of Sanskrit Shabdmitra.

Level-wise interface

Level-Wise interface is designed for non-formal setup where any individual can learn Sanskrit depending upon his/her prior knowledge and language acquisition capabilities. In this scenario, Sanskrit Shabdmitra is focused on self-learning, which is as per the convenience of an individual. However, we should take into the account the nature of mother tongue (L1) and second language (L2) acquisition. Figure 6 shows an interface of Sanskrit Shabdmitra.

The level-wise interface is a big challenge as very few people have Sanskrit as their mother tongue. The majority of people study Sanskrit as their second or even third language. Hence, the levels are determined according to the knowledge of an individual. In order to get a better idea of L1 acquisition, researchers have tried to explain how children progress from “no language” or “blank slate” to their mother tongue. Whereas, for L2 acquisition, the process is more complicated as learners already have the knowledge of their mother tongue (Ipek, 2009).

Hence, the level-wise interface is different for first and second language learners. Taking the above scenario into the consideration and taking help from the National Curriculum Framework (NCF)³³ devised by NCERT - Government of India, and Common European Framework of Reference (CEFR)³⁴ by the Council of Europe the following levels for Sanskrit Shabdmitra are determined:

- Novice प्रारम्भिकः (prārambhikah) -
Novice is considered as a basic user where he/she is provided with the basics/fundamentals of language, like, varṇamālā (i.e., Sanskrit alphabet), word formation, etc.
- Intermediate माध्यमिकः (mādhyamikah) -
Intermediate is an independent user who has mastered the basics of Sanskrit and can communicate simple and basic needs. Here, most frequent words are provided.
- Advanced प्रवीणः (pravīṇah) -
Advanced is a proficient user. Here concept meaning with grammatical information is provided.
- Superior विशेषज्ञः (viśeṣajñah) -
Superior is a well versed language user. Here, multiple senses along with their grammatical and lexico-semantic features are provided.

Figure 4 depicts the levels of Sanskrit Shabdmitra.

4.3.2 Features

Sanskrit Shabdmitra has numerous features. Keeping standardization and language education need as a focus, features of Sanskrit Shabdmitra have been designed. In Sanskrit Shabdmitra, there are tool specific features and lexico-semantic features. Tool specific features are designed considering the usability and accessibility of the tool while teaching and learning Sanskrit. Lexico-semantic features are features which are specific to the word in picture. Lexico-semantic features are given in tables 2 and 3, there are two wide sections of features, viz., ‘Derived features’, which are derived from Sanskrit Wordnet and ‘Advanced features’, which are additional features specially designed considering the properties of Sanskrit language along with the interest of various stakeholders of this digital aid. Sanskrit Wordnet does not provide

³³<http://www.ncert.nic.in/rightside/links/pdf/framework/english/nf2005.pdf>

³⁴<https://www.babbel.com/en/magazine/how-and-why-to-determine-language-proficiency/>

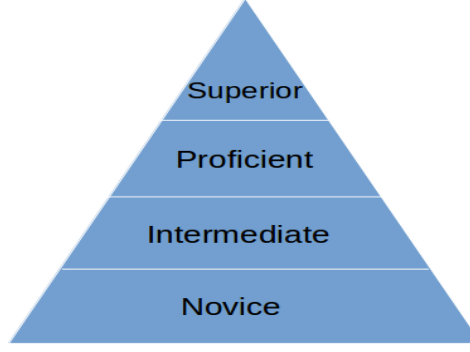


Figure 4: User levels of Sanskrit Shabdmitra

संस्कृत शब्दमित्र

गृहम्

CBSE

Class 6

Lesson 2

अपरः

दृष्टया

खादित्यामि

अन्यः

अकथयत्

गृहं

नीत्वा

भक्षयिष्यामि

इति

बादल

नाम

प्रातिपदिकम् - गृह

परिभाषा - मनुष्यैः इष्टिकादिभिः विनिर्मितं वासस्थानम् ।

प्रयोगः उद्धरणं वा - गृहिण्या एव गृहं शोभते ।

लिङ्गम् - नपुंसकलिङ्गम्

समानार्थी शब्दः - गेहम्, सदनम्, भवनम्, आलयः

विभक्तिः - द्वितीया

वचनम् - एकवचनम्

Figure 5: Class-Wise Interface for Sanskrit Shabdmitra

morphological features, however, Sanskrit Shabdmitra provides them. Table 2 shows the Derived features and Table 3 shows the Advanced features of Sanskrit Shabdmitra. These features rendered along with input word (search word) in interface of the Sanskrit Shabdmitra. Following are the details of these features.

Tool Specific Features

- **Standardization:** Standardization is a unique feature of Shabdmitra wherein all Shabdmitra of all Indian languages are interlinked. This inter-linkage is established using a unique identifier of a synset, called as a synset id. This feature has been inherited from IndoWordNet in which different wordnets are interlinked on the basis of synset id. Hence someone who is learning Hindi can see Sanskrit word for the same concept. Similarly, common Sanskrit words in Hindi for e.g., animals, numbers, flowers, body parts, etc. are unique across all the languages. This way we can attain standardization. Under standardization, we can separate synsets as per the classification of synsets as shown in Table 1; Similarly, illustrations can be shared across all the Indian languages.
- **Varṇamālā:** Sanskrit varṇamālā (alphabets) in Devanagari form is made available in the interface. Here, each of the letter of varṇamālā is displayed in animated form. This can help a learner in understanding the pattern of alphabet writing. Also, pronunciation of the

संस्कृत शब्दमित्र

घर

Levels

प्रारम्भिक: (Novice)

> माध्यमिक: (Intermediate)

प्रवीण: (Advanced)

विशेषज्ञ: (Superior)

अक्ष:



नाम

प्रातिपदिकम् - अक्ष

परिभाषा - आयताकृतिघनः येन द्यूतकाराः दीव्यन्ति।

प्रयोगः उद्धरणं वा - मोहनः अक्षैर्द्वने पटुः ।

लिङ्गम् - पुल्लिङ्गम्

समानार्थी शब्दः - पाशः , पाशकः

Figure 6: Level-Wise Interface for Sanskrit Shabdmitra

same is provided separately in the interface.

- Picture depiction: In Sanskrit Wordnet, there are several concepts which are difficult to explain using the gloss itself. For example, the concept of चषकः (caṣakaḥ, a glass) in Sanskrit is explained as - कषायादिपानार्थम् उपयुक्तं मृद्धात्वादिभिः विनिर्मितं पात्रम्। (kaṣāyādipānārtham upayuktaṁ mṛd-dhātṵvādibhiḥ vinirmitaṁ pātram, a container for holding liquids while drinking).



Figure 7: Picture depicting the concept चषकः (caṣakaḥ, a glass)

This gloss seems to be difficult for lower level learners to understand the concept due to the presence of some difficult words. However, as shown in figure 7, this can be easily understood with the help of a picture. Hence, pictures and illustrations help in differentiating the fine-grained senses found in Wordnet.

- Audio pronunciation: Shabdmitra interface has two types of audio pronunciation viz., मन्दम् (mandam, slow) and सामान्यम् (sāmānyam, normal). The slow-paced pronunciation provides the syllable-based output wherein each syllable is pronounced slowly, one at a time. This helps in understanding the sound structure of a syllable. Whereas for the normal

DERIVED FEATURES	
1.	Word (in a synset form):
2.	Original Gloss (परिभाषा)
3.	Original Example (वाक्ये प्रयोगः उद्धरणं वा):
4.	Gender (लिङ्गम्)
5.	Synonyms (समानार्थी शब्दः)
6.	Antonyms (विरुद्धार्थी शब्दः)
7.	Holonymy (अवयवी)
8.	Meronymy (अवयवः)
9.	Hypernymy (पराजातिः)
10.	Hyponymy (अपराजातिः)

Table 2: Derived Features of Sanskrit Shabdmitra

paced pronunciation, the words are pronounced at a normal pace. These audio features provided with Shabdmitra help in understanding the pronunciation and getting audio clarity of a word.

Derived Features

- Word (in a synset form) - The word which is stored and available in Sanskrit Wordnet synset is shown in this field.
- Original Gloss (परिभाषा) - If gloss is simple enough to understand then the original Sanskrit Wordnet gloss having same sense and synset id is kept as it is and rendered in this field, else a simplified gloss is rendered (This has been explained in the section 'Simplified Gloss' below).
- Original Example (वाक्ये प्रयोगः उद्धरणं वा) - Similarly, by default the original example sentence is retained.
- Gender (लिङ्गम्) - A gender of the word is directly taken from the Sanskrit Wordnet database.
- Synonyms (समानार्थी शब्दः) - Most frequent synonymous words of input word are displayed here. Right now the tool allows to display maximum of 5 words.
- Antonyms (विरुद्धार्थी शब्दः) - Antonyms of input word are displayed in this field.
- Holonymy (अवयवी) - A semantic relation that holds between a whole and its parts.
- Meronymy (अवयवः) - Relation between lexical units where the objects, etc., denoted by one are parts of those denoted by other.
- Hypernymy (पराजातिः) - A semantic relation between two synsets to capture super-set hood.
- Hyponymy (अपराजातिः) - A semantic relation between two synsets to capture sub-set hood.

Advanced Features

- Word (inflected form) - This particular feature is specific to a class-wise interface wherein an input word (i.e., word appeared in the textbook) which is having an inflected form is displayed.
- Word [in root form] (प्रातिपदिकम्) - This is applicable only to nouns which are in nominative singular form. Here, root word of the noun is displayed.

ADVANCED FEATURES	
1.	Word (inflected form)
2.	Word [in root form] (प्रातिपदिकम्)
3.	Simplified Gloss (परिभाषा)
4.	Simplified Example (वाक्ये प्रयोगः उद्धरणं वा)
5.	Type of Noun (संज्ञायाः प्रकारः)
6.	Type of Adjective (विशेषणस्य प्रकारः)
7.	Type of Verb (क्रियायाः प्रकारः)
8.	Type of Adverb (क्रियाविशेषणस्य प्रकारः)
9.	Case (विभक्तिः)
10.	Lakāra (लकारः)
11.	Person (पुरुषः)
12.	Number (वचनम्)
13.	Affix, Suffix (प्रत्ययः)
14.	Preposition, Prefix (उपसर्गः)
15.	Accent (स्वरः)
16.	Dhātuprakārah (धातुप्रकारः)
17.	Gaṇah (गणः)
18.	Padam (पदम्)
19.	With Augment 'इ'
20.	Transitivity (कर्मकत्वम्)

Table 3: Advanced Features of Sanskrit Shabdmitra (feature numbers 1, 2, and 9 to 19 are morphological features)

- Simplified Gloss (परिभाषा) - Concepts which are difficult to understand are simplified.
For example, in SWN for a word 'अक्षः' (akṣaḥ) the original gloss is 'काष्ठस्य वा अस्थिनः आयताकृतिघनः येन द्यूतकाराः दीव्यन्ति' (kāṣṭhasya vā asthinaḥ āyataakṛtighanaḥ yena dyūtakārah dīvyanti, a cubical shaped piece made of wood or bone used by gamblers for playing). Such a gloss, being too elaborate and difficult to follow at the beginner's level, has been simplified to: आयताकृतिघनः येन द्यूतकाराः दीव्यन्ति (āyataakṛtighanaḥ yena dyūtakārah dīvyanti, a cubical shaped piece used by gamblers for playing).
- Simplified Example (वाक्ये प्रयोगः उद्धरणं वा) - Similarly, examples are simplified.
- Type of Noun (संज्ञायाः प्रकारः) - If the input word is a noun then it is assigned with the prescribed types of nouns. This information is usually taken from ontological database of IndoWordNet.
- Type of Adjective (विशेषणस्य प्रकारः) - If the input word is an adjective then it is assigned with the prescribed types of adjectives. This information is usually taken from ontological database of IndoWordNet.
- Type of Verb (क्रियायाः प्रकारः) - If the input word is a verb then it is assigned with the prescribed types of verbs. This information is usually taken from ontological database of IndoWordNet.
- Type of Adverb (क्रियाविशेषणस्य प्रकारः) - If the input word is an adverb then it is assigned with the prescribed types of adverbs. This information is usually taken from ontological database of IndoWordNet.

- Countability (गणनीयता) - Nouns can be either countable or uncountable. Accordingly, the countability is assigned to the nouns. Countable nouns are those that refer to something that can be counted. On the other hand nouns which do not typically refer to things that can be counted, are Uncountable nouns³⁵.
- Case (विभक्तिः) - The input word can belong to any of the eight cases. They are listed as below:
 - Nominative - प्रथमा (prathamā)
 - Accusative - द्वितीया (dvitīyā)
 - Instrumental - तृतीया (tṛtīyā)
 - Dative - चतुर्थी (caturthī)
 - Ablative - पञ्चमी (pañcamī)
 - Genitive - षष्ठी (ṣaṣṭhī)
 - Locative - सप्तमी (saptamī)
 - Vocative - संबोधन (sambodhana)
- Lakāra (लकारः) - This is verb specific property of a word which helps in identifying the tense, aspect and modality of a word. The input word can belong to any of the 10 types of lakāra. They are listed as below:
 - laṭ - लट् (laṭ)
 - lañ - लङ् (lañ)
 - loṭ - लोट् (loṭ)
 - vidhiliñ - विधिलिङ् (vidhiliñ)
 - āśīrliñ - आशीर्लिङ् (āśīrliñ)
 - liṭ - लिट् (liṭ)
 - luṭ - लुट् (luṭ)
 - luñ - लुङ् (luñ)
 - lṛṭ - लृट् (lṛṭ)
 - lṛñ - लृङ् (lṛñ)
- Person (पुरुषः) - This is a verb specific property of a word wherein the verb can appear in the sense of person viz. the first (उत्तमः), second (मध्यमः) and third (प्रथमः) (Pāṇini and Vasu, 1962) [1.4.101]
- Number (वचनम्) - Inflectional category basically distinguishing reference to one individual from reference to more than one.
- Affix, Suffix (प्रत्ययः) - There are six main kinds of affixes given in Sanskrit grammar viz., सुप्, तिङ्, कृत्, तद्धित, धातुप्रत्ययः [(i.e. सन्, क्यप्, etc.)] and स्त्रीप्रत्ययः. Right now, in Sanskrit Shabdmitra only first 3 types of affixes i.e. सुप्, तिङ्, कृत् are shown.
- Preposition, Prefix (उपसर्गः) - The word उपसर्ग originally meant only 'a prefixed word'. These prefixes are always used along with a verb (Abhyaṅkara and Shukla, 1977) [pg 88]
- Accent (स्वरः) - This property is possessed only by vowels and not by consonants (Abhyaṅkara and Shukla, 1977) [pg 438]. Accents are basically found in vedic texts. Except traditional schools, vedic texts are not part of the school syllabus viz., CBSE, ICSE, etc. Hence, accents are not introduced in primary level of Sanskrit Shabdmitra. However, words with accents shall be introduced in advanced levels. Following are the types of accents:

³⁵<https://www.lexico.com>

- उदात्तः the acute accent defined by Panini (Pāṇini and Vasu, 1962) [1.2.29]. The acute is the prominent accent in a word (Abhyaṅkara and Shukla, 1977) [pg 81]. According to the position in the word, the acute accent has following sub-types:
 - * आद्युदात्तः a word beginning with an acute accent i.e. which has got the first vowel accented acute.
 - * मध्योदात्तः the acute accent to the middle vowel which is neither the initial nor the final.
 - * अन्तोदात्तः a word with its last vowel accented acute.
- अनुदात्तः the grave accent defined by Panini (Pāṇini and Vasu, 1962) [1.2.30].
- स्वरितः the circumflex accent defined by Panini (Pāṇini and Vasu, 1962) [1.2.31].
- Dhātuprakārah (धातुप्रकारः) - There are different types of root verb as follows:
 - औपदेशिकधातुः (पाणिनीयधातुपाठे उपदिष्टः) Panini has given a long list of roots under ten groups named as औपदेशिकधातुः or primary roots.
 - आदेशिकधातुः (सनाद्यन्तादिधातवः). There are two types of them, they are as follows:
 - * roots derived from roots. These are classified into three types:
 - causative (णिजन्त)
 - desiderative (सन्नन्त)
 - intensive (यङन्त)
 - * roots derived from nouns.
 - वैदिकधातुः roots found in vedic literature.
 - सौत्रधातुः roots mentioned specifically in paninian rule only.
- Gaṇah (गणः) - There is a long list of roots under the following ten groups. They are as follows:
 - भ्वादिगणः (bhvādigaṇah)
 - अदादिगणः (adādigaṇah)
 - जुहोत्यादिगणः (juhotyādigaṇah)
 - दिवादिगणः (divādigaṇah)
 - स्वादिगणः (svādigaṇah)
 - तुदादिगणः (tudādigaṇah)
 - रुधादिगणः (rudhādigaṇah)
 - तनादिगणः (tanādigaṇah)
 - क्र्यादिगणः (kryādigaṇah)
 - चुरादिगणः (curādigaṇah)
- Padam (पदम्) - A technical term for the affixes. There are three types of padam:
 - parasmaipadam परस्मैपदम् term used in grammar with reference to the personal affixes ति (ti), तः (ta), etc.
 - ātmanepadam आत्मनेपदम् a technical term for the affixes त (ta), आताम् (ātām), etc.
 - ubhayapadam उभयपदम् a technical term in which a specific group of verbs are from both parasmaipada and aatmanepada (Abhyaṅkara and Shukla, 1977) [pg 92]
- With Augment 'इ' - Here इ (i) is prefixed in the case of root.
 - अनिट् (aniT) roots अनिट् does not allow the augment इ to be prefixed.
 - सेट् roots सेट् always allows the augment इट् to be prefixed.
 - वेट् roots optionally admit the application of the augment इ.

CBSC Syllabus	All words	Unique words
Class VI	1499	813
Class VII	2655	1604
Class VIII	3072	1987
Class IX	2701	1814
Class X	2902	1989
Total	12829	8207
All class unique words		6784

Table 4: Sanskrit word-collection statistics

- Transitivity (कर्मकत्वम्) - karmakatvam can be one among the two as follows:
 - सकर्मकः sakarmakah, transitive
 - अकर्मकः akarmakah, intransitive

4.3.3 Shabdmitra Enriches Wordnet

It is noticed that the development of Sanskrit Shabdmitra leads to the enrichment of SWN. It is a two-way process in which SWN helps Sanskrit Shabdmitra by providing the resource, while Sanskrit Shabdmitra helps SWN by providing additional words and properties, hence enriching the same. Table 4 depicts the count of words which are collected and unique words from classes VI to X under the CBSE board.

4.3.4 Applications

There are various applications of Sanskrit Shabdmitra. Some of them are listed as below:

- Sanskrit Shabdmitra is an educational tool for teaching and learning Sanskrit vocabulary.
- It also acts as a teaching and learning aid for teachers in school setup.
- It can also be used for testing the Sanskrit language knowledge of an individual.
- This tool can be of great help for conducting and preparing Sanskrit competitive exams.
- It can be used to explain tatsama and tadbhava words in other languages.

5 Conclusion

In this paper, how Sanskrit Wordnet can be used for developing educational application has been explained. It is also demonstrated how a semantically rich lexical resource like Wordnet, originally developed for research purpose can be remodeled for practical usage in education domain.

Sanskrit Shabdmitra is one such comprehensive e-learning aid which helps in learning Sanskrit language, pronunciation, grammar and understanding the concepts through images, definition and examples. It caters to a wider range of audience ranging from school children to individual learners at different levels, i.e., from novice to the superior. The tool, Sanskrit Shabdmitra presented here is a multi-modal, multi-layered Sanskrit language teaching and learning aid which can be used for formal and informal learning environments. Further, Shabdmitra acts as a platform, as a resource as well as a brand. It helps in enriching the Sanskrit Wordnet and vice versa.

6 Future Work

In Future, we plan to incorporate question answering system which can help in understanding the knowledge of the user, also which can help in understanding the level at which he can start learning Sanskrit. Also, the tool will be improved with the inclusion of gamification, bilingual as well as multilingual learning and teaching under Shabdmitra platform.

Acknowledgements

We would like to thank the entire Hindi Shabdmitra Team; Center for Indian Language Technology (CFILT), and Indian Institute of Technology Bombay (IITB) for providing us necessary resource and support. We further thank TCTD, IIT Bombay³⁶ for providing the necessary support. Finally, we thank reviewers for their comments, positive remarks and encouragements.

References

- K.V. Abhyankara and J.M. Shukla. 1977. A Dictionary of Sanskrit Grammar. Number no. 134 in A dictionary of Sanskrit grammar. Oriental Institute.
- Pushpak Bhattacharyya. 2010. Indowordnet. In Proceedings of Lexical Resources Engineering Conference (LREC), Malta.
- Pushpak Bhattacharyya. 2017. Indowordnet. In *The WordNet in Indian Languages*, pages 1–18. Springer.
- Sudha Bhingardive and Pushpak Bhattacharyya. 2017. Word sense disambiguation using indowordnet. In *The WordNet in Indian Languages*, pages 243–260. Springer.
- Sudha Bhingardive, Tanuja Ajothkar, Irawati Kulkarni, Malhar Kulkarni, and Pushpak Bhattacharyya. 2014. Semi-automatic extension of sanskrit wordnet using bilingual dictionary. In Proceedings of the Seventh Global Wordnet Conference, pages 324–329.
- Sudha Bhingardive, Hanumant Redkar, Prateek Sappadla, Dharendra Singh, and Pushpak Bhattacharyya. 2016. Indowordnet:: Similarity computing semantic similarity and relatedness using indowordnet. In *Global WordNet Conference*, page 39.
- Heidi Brumbaugh. 2015. Self-assigned ranking of L2 vocabulary: using the Bricklayer computer game to assess depth of word knowledge. Ph.D. thesis, Arts & Social Sciences:.
- Amit C. Hiray. 2015. Teaching and Learning of EAP Vocabulary: A Web-based Integrative Approach at the Tertiary Level in India. Ph.D. thesis, Dept. of HSS, IIT Bombay.
- X Hu, AC Graesser, Tutoring Research Group, et al. 1998. Using wordnet and latent semantic analysis to evaluate the conversational contributions of learners in the tutorial dialog. In Proceedings of the international conference on computers in education, volume 2, pages 337–341.
- Hulya Ipek. 2009. Comparing and contrasting first and second language acquisition: Implications for language teachers. *English Language Teaching*, 2(2):155–163.
- Malhar Kulkarni, Chaitali Dangarikar, Irawati Kulkarni, Abhishek Nanda, and Pushpak Bhattacharyya. 2010a. Introducing sanskrit wordnet. In *Principles, Construction and Application of Multilingual WordNets*, Proceedings of the 5th GWC, edited by Pushpak Bhattacharyya, Christiane Fellbaum and Piek Vossen, Narosa, page 257–294. Narosa Publishing House, New Delhi.
- Malhar Kulkarni, Irawati Kulkarni, Chaitali Dangarikar, and Pushpak Bhattacharyya. 2010b. Gloss in sanskrit wordnet. In Proceedings of Sanskrit Computational Linguistics, pages 190–197. Berlin: Springer-Verlag / Heidelberg.
- Malhar Kulkarni. 2017. Sanskrit wordnet at indian institute of technology (iitb) mumbai. In *The WordNet in Indian Languages*, pages 231–241. Springer.

³⁶<http://www.tatacentre.iitb.ac.in/>

- Chih-Cheng Lin. 1997. Semantic network for vocabulary teaching. *Journal of National Taiwan Normal University*, (42):43–54.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Roberto Navigli and Paola Velardi. 2004. Learning domain ontologies from document warehouses and dedicated web sites. *Computational Linguistics*, 30(2):151–179.
- Pāṇini and Srisa Chandra Vasu. 1962. *The Aṣṭādhyāyī of Pāṇini*. Motilal Banarsidass.
- Hanumant Redkar, Jai Paranjape, Nilesh Joshi, Irawati Kulkarni, Malhar Kulkarni, and Pushpak Bhattacharyya. 2014. Introduction to synskarta: An online interface for synset creation with special reference to sanskrit. In *11th International Conference on Natural Language Processing (ICON-2014)*, Goa, India, page 229.
- Hanumant Redkar, Nilesh Joshi, Sandhya Singh, Irawati Kulkarni, Malhar Kulkarni, and Pushpak Bhattacharyya. 2016. *Samāsa-kartā: An online tool for producing compound words using indowordnet*. In *8th Global WordNet Conference*.
- Hanumant Redkar, Sandhya Singh, Meenakshi Somasundaram, Dhara Gorasia, Malhar Kulkarni, and Pushpak Bhattacharyya. 2017a. Hindi shabdmitra: A wordnet based e-learning tool for language learning and teaching. In *Proceedings of the 4th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2017)*, pages 23–28, Taipei, Taiwan, December. Asian Federation of Natural Language Processing.
- Hanumant Redkar, Nilesh Joshi, Sayali Khare, Lata Popale, Malhar Kulkarni, and Pushpak Bhattacharyya. 2017b. Hindi shabdmitra: A wordnet based tool for enhancing teaching-learning process. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON 2017)*, Jadavpur University, Kolkata, India, December.
- Koun-Tem Sun, Huang Yueh-Min, and Liu Ming-Chi. 2011. A wordnet-based near-synonyms and similar-looking word learning system. *Journal of Educational Technology & Society*, 14(1):121.