

Vaijayantīkośa Knowledge-Net

Aruna Vayuvegula*, Satish Kanugovi†, Sivaja S Nair‡, Shivani V§ and Mahalakshmi¶

Karnataka Samskrit University

Bangalore, India

33aruna@gmail.com, satishk.rao@gmail.com,

sivaja.s.nair@gmail.com, shivani.ksu@gmail.com and prasmx@gmail.com

Abstract

A kośa (lexicon) is a literary work that provides a comprehensive understanding of words by arranging them along with their synonyms and other words that are semantically related. Its format has been designed to include not just ontological classification, but to give a holistic idea of a concept represented by the word. This allows a thorough understanding of the words, and also the knowledge they embody. Vaijayantīkośa is a popular Sanskrit lexicon that contains words from spoken language as well those used in Vedic literature. To facilitate dissemination of this knowledge, a web-based tool, Vaijayantīkośa Knowledge Net, is created for easy access and analysis of the words in the kośa. The objective of the tool is to provide information to researchers from different fields of study to explore the knowledge contained in the kośa with the help of synsets and ontological structure.

Key words: Vaijayantīkośa, Synset, Ontology, KnowledgeNet, Semantic relations

1 Introduction

Sanskrit is rich with domain-specific and subject-specific kośa literature. They are written in verse format enabling them to be memorized easily by students. Generally kośa, in the Indian tradition of knowledge representation, is a grouping of words with semantic relations to provide comprehensive understanding of the word and its ontological classification. The ontological classification and knowledge structure in Sanskrit kośas have been described in detail by Kulkarni, A (2010).

Patkar (1981) in his book “History of Sanskrit Lexicography”, lists at least 81 lexicons that were written in Sanskrit between 400 BC and 1800 AD. Vogel (1975) in his work, ‘Indian Lexicography’ details the characterization of Indian lexica and lists down over forty unique dictionaries, many special, bilingual and multilingual dictionaries. Unfortunately, many of these works have been lost and we are left with very few of these treasures. Hence, there is a need to ensure that the existing lexicons are well-preserved for posterity and technology can be a great asset to achieve this goal.

Amarakośa is the most authoritative and ancient thesaurus of Sanskrit. There have been several commentaries and translations of the lexicon (Patkar, 1981, pp. 19-21) both in Indian as well as foreign languages. In recent times, it has also captured the interest of computational linguists. Nair, in her PhD thesis (Nair, 2011) has detailed the knowledge structure of Amarakośa and developed the tool, Amarakośa Knowledge Net (AKN), that systematically represents the links between words based on a structured table in a dynamic manner. She has suggested in her thesis that AKN can serve as a model for developing tools for other kośas.

As part of the Post-graduate diploma in Sanskrit computational linguistics program, we take this suggestion forward and develop the Vaijayantīkośa Knowledge Net (VKN) tool to capture

the knowledge structure of Vaijayantīkośa. The paper introduces Vaijayantīkośa and details the different aspects of the VKN development in the following sections.

2 Vaijayantīkośa (VK)

VK is written by Yādavaprakāśa between 10th and 11th century (Bühler, 1887). He lived in the southern part of India, near the present day Kanchipuram in Tamilnadu (Oppert, 1893, p. 2). VK not only has a rich vocabulary of words for common usage, it also has a large number of terms from the Vedas.

Though there are many manuscripts on VK, in different Indian languages, none of them is complete, except one manuscript in Malayalam language. (Oppert, 1893, pp. 3-4).

For the purpose of this work, we have referred to the following two texts of VK.

1. The Vaijayantī of Yādavaprakāśa compiled by Gustav Oppert

This version has introduction by Gustav in English and an elaborate section of vocabulary with meanings in English. Gustav has painstakingly referred to 11 manuscripts and consolidated all the kāṇḍas as one entity (Oppert, 1893).

2. Vaijayantīkośa compiled by Sri. Pandit Haragovindashastry.

This version has introduction by Pandit Haragovindashastry in Hindi and appears, to a large extent, based on Gustav Oppert's work itself. He gives a brief commentary on the uniqueness of the lexicon and adds glossary of words at the end with references to the ślokas where the words appear (Haragovindashastry, 1971).

Bühler (1887) gives an overview of VK, its structure and information about its author. Kulkarni refers to VK while giving an overview of lexicographic traditions in India and Sanskrit (Kulkarni, 2010). Kaur also touched upon VK through a taxonomical analysis of early Sanskrit literature (Kaur & Singh, 2018). Vogel touches upon VK while chronicling Indian Lexicography and gives brief details about the style and classification adopted by the author (Vogel, 1975). Some regional scholars have also referred to VK in their works. For example, Mallinatha, in *Amarapadapārijāta* (commentary on Amarakośa) provides close to 212 citations from VK (Nair, 2011).

For this project, the compilation of VK by Gustav Oppert has been taken because of the comprehensiveness of his work as well as the detailed vocabulary of words with meanings in English.

2.1 Structure of VK

The author, Yādavaprakāśa has arranged the words into kāṇḍas and adhyāyas based on a clear ontological structure. The kāṇḍas are named according to the major topic covered. For example, the antarīkṣakāṇḍa consists of all the words related to the sky, universe, astronomy, astrology etc.

Each kāṇḍa is further divided into adhyāyas with semantically related words, arranged together according to context, in the form of ślokas. The classification is detailed in the Figure 1.

1. VK consists of nearly 20000 entries of words listed in verse form.
2. It begins with a maṅgalaśloka followed by nine and a half verses of paribhāṣaślokas which provide pointers to decode the gender information of the words.

स्त्रीपुंनपुंसकं लिङ्गं सङ्कीर्णं तच्च पञ्चधा।

नृस्त्री नृषण्डषण्डस्त्री त्रिलिङ्गं वाच्यलिङ्गकम्॥१.१.३॥¹

समासे स्युः पृथक् सर्वे शब्दा बहुवचोऽन्तके॥१.१.५॥

More rules for interpreting the liṅga (gender) of the words are described in 58 ślokas of Liṅgasaṅgrahādhya (of Śeṣakāṇḍa).

3. There are two major divisions of the kośa - Paryāyabhāga (synonymous words) and Nānārthabhāga (polysemous words).
4. There are five kāṇḍas under Paryāyabhāga and three under Nānārthabhāga.
5. The kāṇḍas are further divided into adhyāyas; they are 43 in total.
6. The structure of VK is represented below (Figure 1).

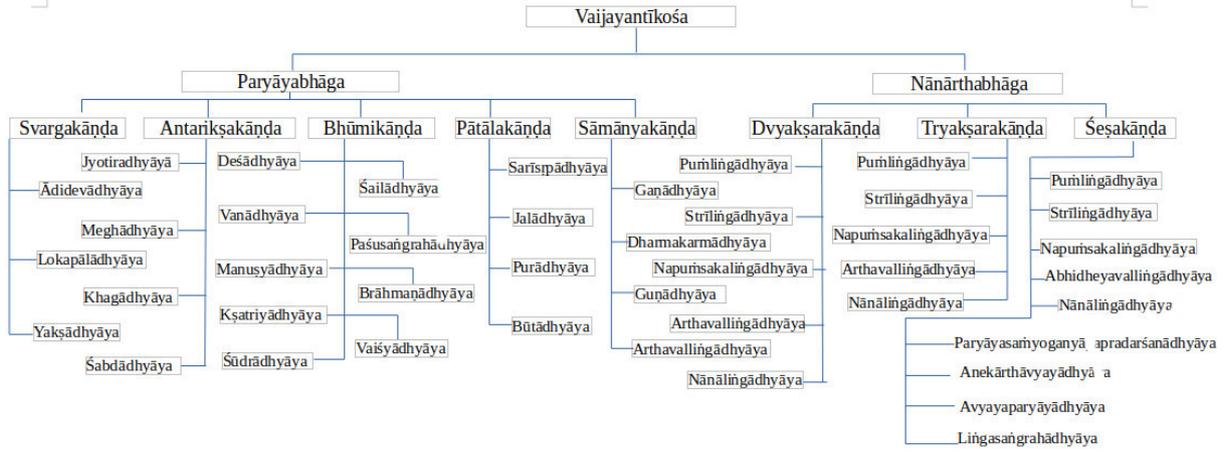


Figure 1: Classification of VK

7. Ślokas in VK contain words, their synonyms and meanings. In some cases, probably where the author found it necessary, information pertaining to gender, brief description of the term may also be included.
8. VK emphasizes understanding a concept at greater depth and precision.

2.1.1 Semantic Arrangement of Words in VK

In VK, the kāṇḍas are arranged based on a particular theme. Kāṇḍas are further divided into adhyāyas which are based on sub-themes. Adhyāyas contain ślokas that mostly follow semantic order with occasional violations. Ślokas contain words that are related to a concept. A given word is typically followed with its synonyms and subsequently other relations, like पति-पत्नीभावः (husband – wife relation), जन्य-जनकभावः (child – parent relation), स्व-स्वामिभावः (owner – property relation), सेव्य-सेवकभावः (lord - servant relation), धर्म-धर्मिभावः (property - locus relation), गुण-गुणिभावः (quality - qualifier relation) etc. For example, in concept Viṣṇu, first 53 words form a synset. Subsequently, the author lists words that refer to powers of Viṣṇu. They are followed by possessions of Viṣṇu and so on. Nevertheless, there is a pattern that perhaps reflects the logic of the times it was written.

Given below is the example of the word how Viṣṇu is dealt in VK.

¹Śloka reference: The position of a śloka in the VK is represented numerically as x.y.z, where x=adhyāya number, y = kāṇḍa number in the adhyāya, and z = the śloka number in the kāṇḍa. For example in this śloka 1.1.3.

Example 1: Concept of विष्णुः

The śloka 1.1.10 to 1.1.38 from ādidevādhyāya of svargakāṇḍa describe the concept of Viṣṇu with different relations. See Figure 2.

field 1 : word in sanskrit, field 2 : English equivalent in (), field 3 : number in synset in (), field 4 : kāṇḍa.adhyāya.śloka in ()

- विष्णुः (epithet of Viṣṇu)(53)(1.1.10 - 1.1.15)
वैष्णवी (power of Viṣṇu)(9)(1.1.16)
कौस्तुभः (jewel of Viṣṇu)(1)(1.1.17)
श्रीवत्सः (mark on Viṣṇu)(1)(1.1.17)
नन्दकः (sword of Viṣṇu)(1)(1.1.17)
शार्ङ्गः (bow of Viṣṇu)(1)(1.1.17)
पाञ्चजन्यम् (conch of Viṣṇu)(1)(1.1.17)
सुदर्शनम् (discus of Viṣṇu)(1)(1.1.17)
कौमोदकी (mace of Viṣṇu)(1)(1.1.18)
नरसिंहः (incarnation of Viṣṇu)(1)(1.1.18)
वामनः (incarnation of Viṣṇu)(10)(1.1.19 - 1.1.20)
परशुरामः (incarnation of Viṣṇu)(2)(1.1.20)
श्रीरामः (incarnation of Viṣṇu)(15)(1.1.20 - 1.1.24)
बलभद्रः (incarnation of Viṣṇu)(20)(1.1.22 - 1.1.24)
संवर्तकम् (Plough of Balabhadra)(1)(1.1.24)
सौनन्दनम् (pestle of Balabhadra)(1)(1.1.25)
कृष्णः (incarnation of Viṣṇu)(10)(1.1.25 - 1.1.26)
दारुकः (Charioteer of Kriṣṇa)(1)(1.1.26)
वसुदेवः (father of Kriṣṇa)(3)(1.1.26)
मन्मथः (god of love, son of Viṣṇu)(25)(1.1.27 - 1.1.29)
अनिरुद्धः (son of Manmatha)(3)(1.1.29)
नरनारायणः (incarnation of Viṣṇu)(2)(1.1.30)
हयग्रीवः (incarnation of Viṣṇu)(2)(1.1.30)
आदिशेषः (incarnation of Viṣṇu)(1)(1.1.30)
व्यासः (incarnation of Viṣṇu)(6)(1.1.30)
दत्तात्रेयः (incarnation of Viṣṇu)(1)(1.1.31)
कल्किः (incarnation of Viṣṇu)(1)(1.1.31)
कपिलः (incarnation of Viṣṇu)(3)(1.1.31)
व्यासः (incarnation of Viṣṇu)(6)(1.1.31- 1.1.32)
बुद्धः (incarnation of Viṣṇu)(32)(1.1.32 - 1.1.35)
लक्ष्मीः (wife of Viṣṇu)(10)(1.1.36)
गरुडः (vehicle of Viṣṇu)(12)(1.1.37 - 1.1.38)

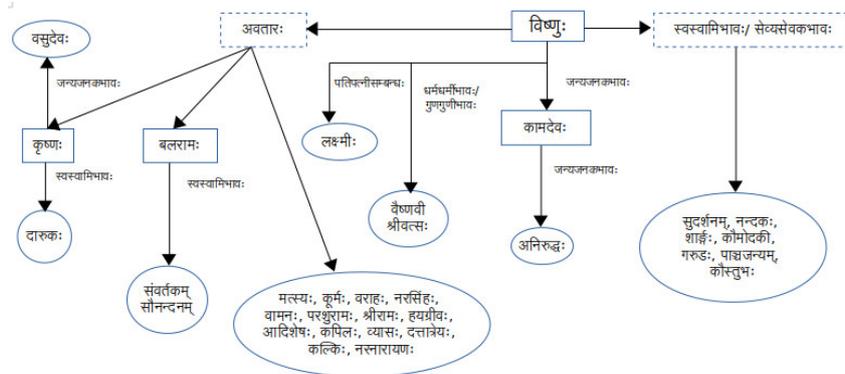


Figure 2: Relations of Viṣṇu

Example 2: Concept of कालः

In VK, the reference to kāla is from śloka 2.1.52 to 2.1.54, which is a total of 43 words in the jyotiradhyāya of antarikṣakāṇḍa. The concept of kāla starts with the smallest unit of time which is referred to as तुटिः (moment). Subsequently, higher units of time are mentioned as depicted below:

कालः (time)(3)(2.1.52)

तुटिः (moment)(2)(2.1.52)

लघ्वक्षरकः (space of two moments)(1)(2.1.52)

अक्षरपातकः (space of two laghvakṣarakas)(1)(2.1.52)

निमेषः (space of two akṣarapātakas)(1)(2.1.53)

लिप्तिका (space of two nimeṣas)(1)(2.1.53)

काष्ठा (space of nine liptikās)(1)(2.1.53)

लवः (space of two kāṣṭhas)(1)(2.1.53)

कला (space of five lavas)(1)(2.1.53)

लेशः (space of twelve kalās)(1)(2.1.54)

क्षणः (space of 16 leśas)(1)(2.1.54)

नाडी (space of six kṣaṇas)(1)(2.1.54)

मुहूर्तः (space of twelve nādis)(1)(2.1.54)

घटिका (space equal to one muhūrta)(2)(2.1.54)

Here we can see the hierarchical order of the words which is connected through the relation अवयव-अवयविसम्बन्धः. Subsequent ślokas i.e. 2.1.55 to 2.1.73 also deal with the concept of kāla but has not been depicted here due to lack of space. A few observations on examining the concept of the word kāla are as follows:

- A very logical and precise structure of division of time has been adopted starting from the lowest measure of time.
- A very systematic division of time until it spans 24 hours or one day is seen. Then, there is the first violation of nesting where day is followed by night and the author goes on to describe night, different kinds of night. Within the nesting of night too, after describing different kinds of night, he suddenly introduces darkness and then goes on to describe different types of darkness.
- After this, there is the third violation of nesting when he goes back to day and then defines different parts of the day followed by different parts of night. Next, he picks terms that talk about space of three hours (which is relevant to both day and night), lucky portion of the day, dawn and twilight. He then ends by addressing a lunar day and different days in a lunar month.
- The list is followed by months, seasons, years, yugas etc.

The author often describes the qualities of a particular term. For example, under the main word 'sun', the term sunray is given. The author lists down 22 words under the concept of sunrays. These words do not appear to be synonymous but indicate a more complex idea that needs further research.

तासां शतानि चत्वारि रश्मीनां वृष्टिसर्जने।
शतत्रयं हिमोत्सर्गे तावत्घर्मस्य सर्जने॥ २.१.१७²

3 Vaijayantikośa Knowledge Net (VKN)

VKN is a web-based tool to access knowledge embodied in VK by providing comprehensive information related to the word including meanings, synonyms and relations with other words.

²This śloka is only a small extract of the group of verses that are referred under sunrays.

3.1 Scope of the present project

VK is a voluminous lexicon with approximately 20,000 entries of words. However, for developing this version of the web-tool, the first two kāṇḍas mainly the svargakāṇḍa and antarīkṣakāṇḍa have been taken, which contain 3,000 entries. The output of the web-tool is the synset and the set of related words of a given input - padam (word in its first person singular form) or prātipadikam (stem). The tool consciously confined to the first 3,000 entries as new fields and features kept evolving through the research. For example, including English meanings was not part of the initial plan but was included as it would help users. Once the web-tool is fine-tuned in all respects, it is easier to scale it up to include the entire database.

An Android Application version of the tool is also currently under development. An initial version is available for volunteer testing to get feedback and suggestions on usability. The Android App is briefly described in Section 3.8.

3.2 Data Structure

The first step towards the creation of the web-tool is to digitise the entire kośa. The following categories of information are extracted from the śloka.

प्रातिपदिकम् (stem), पदम् (nominative form), सन्दर्भसूची (reference), लिङ्गम् (gender), अध्यायः (chapter) and काण्डः (section)

For example, the śloka number 47 in the lokapālādhyāya reads as follows:

वातो वायुर्जगत्प्राणश्शुषिलश्श्वसनोऽनिलः।
गन्धवाहो गन्धवहो मातरिक्षा समीरणः॥१.२.४७॥

The words are extracted and categorized as in Table 1.

प्रातिपदिकम्	पदम्	सन्दर्भसूची	लिङ्गम्	अध्यायः	काण्डः	आङ्ग्लार्थः	अर्थः	मुख्यपदम्
वात	वातः	1.2.47.1.1	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
जगत्प्राण	जगत्प्राणः	1.2.47.1.2	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
शुषिल	शुषिलः	1.2.47.1.3	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
श्वसन	श्वसनः	1.2.47.1.4	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
अनिल	अनिलः	1.2.47.1.5	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
गन्धवाह	गन्धवाहः	1.2.47.1.5	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
गन्धवह	गन्धवहः	1.2.47.2.1	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
मातरिक्षन्	मातरिक्षा	1.2.47.2.2	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः
समीरण	समीरणः	1.2.47.2.3	पुं.	लोकपालाध्यायः	स्वर्गकाण्डः	epithet of vāyu	स्पर्शगुणकः पञ्चभूतभेदः	वायुः

Table 1: Information extraction of the synset वायुः

It is to be noted that words from वातः till समीरणः are synonyms, i.e. words with the same meaning.

1. प्रातिपदिकम् is the stem of the tokens from śloka and has been used so that it is compatible with other computational resources such as morphological generator and analyser, various e-lexicons etc.; many of them use प्रातिपदिकम् as input and not the पदम्.
2. पदम् field contains the nominative singular form of the प्रातिपदिकम्, generated using the morphological generator. In the case of नित्यबहुवचनान्त words, the nominative plural form will be taken. If a प्रातिपदिकम् has more than one gender, and masculine form is one of them, the masculine singular form is taken. In case the word has feminine and neuter forms the neuter form of the प्रातिपदिकम्, will be used. These guidelines are based on the

rules of the dictionaries such as Śabdakalpadruma, Vācaspatya etc. This option does not appear in AKN but has been introduced in VKN web-tool to allow users to search for a word using पदम् option in case they are unsure about प्रातिपदिकम् of a particular word. It is hoped that this feature will make the tool user-friendly.

3. सन्दर्भसूची is the reference indicating the precise position of the word in VK using a 5-tuple number as kāṇḍa, adhyāya, śloka, pāda and word number in the pāda. The pāda number and word number in the pāda are entered manually into the database, whereas the other fields are derived automatically.
4. लिङ्गम् - gender information of the word. The gender of a word is decided by the meta-information mentioned by Yādavaprakāśa. Cross reference to लिङ्गसङ्ग्रहाध्याय as well as Oppert's vocabulary is also consulted.
5. अध्यायः refers to the chapter or the adhyāya name to which the entry belongs. The adhyāyas are named based on the topic or subject that the word is categorized under. Thus, this field gives an ontological idea about the word.
6. काण्डः refers to the specific section of VK or kāṇḍa to which the entry belongs.
7. आङ्ग्लार्थः or the meaning in English is an additional field that has been included to document from the translation that Oppert compiles under the vocabulary section of the book. This has been included to ensure VKN is accessible to those who may not be Sanskrit scholars.
8. अर्थः refers to the meaning in Sanskrit given by Yādavaprakāśa in VK. Where ever the meaning is not found in VK, other dictionaries have been referred.
9. मुख्यपदम् or headword represents the synset with synonymous words. Headword is chosen as follows - if the headword used in AK appears in VK synset, that word is chosen as the headword. In case, there is no equivalent word in AK, Oppert's vocabulary at the end of the kośa is referred to choose the headword. There are some challenges in choosing the headword because there are no commentaries on VK that a researcher can refer to in case of doubt. However, effort has been made to ensure that words are chosen as far as possible based on the available resources - Compatibility with AKN and Oppert's vocabulary being a primary guiding forces.

As compared to AKN, three categories, namely - पदम्, meaning in English and meaning in Sanskrit are additional fields incorporated into VKN. The decision to incorporate these additional fields was taken mid-way through the research as it was found to be a useful improvisation over the AKN.

3.3 Relations in VKN

The various relations amongst different headwords are marked in the database. Twelve hierarchical or associative relations are marked in different fields - two kinds of ontological categories, class and attribute are marked in the last two fields. Except ontological categories, all other relations are marked using headwords.

3.3.1 पर्यायवाची (Synset)

The set of words that have similar meaning is defined as a synset. See the example of वातः in table 2. The output synset is displayed in the Figure 6 in the appendix.

3.3.2 अवयव-अवयविभावः (Part-whole Relation)

The अवयव-अवयवि relation is marked to indicate part and whole relation. For example - the synset पक्षः is a part of the synset पक्षी. Each member of the synset पक्षः is related to the members of पक्षी through this relation³.

³See Figure 7. in appendix

3.3.3 परा-अपरासम्बन्धः (Superset-subset Relation)

This field marks परा-अपरासम्बन्धः. For example - the synset मृदुवातः is a kind of वायुः. So the synset मृदुवातः is related to the synset वायुः with परा-अपरा relation. Each member of the synset मृदुवातः is marked to the synset वायुः⁴.

3.3.4 जन्य-जनकभावः (Child-parent Relation)

This field marks जन्य-जनकभावः of two concepts. For example - the synset of पार्वती is related to the synset हिमवान् through जन्य-जनक relation. पार्वती is daughter of हिमवान् and हिमवान् is father of पार्वती.⁵

3.3.5 पति-पत्नीभावः (Husband-wife Relation)

This field is meant for marking पति-पत्नी relation. For example - the synset of शची is related to the synset इन्द्रः with Husband-wife relation. Here इन्द्रः is the husband of शची and शची is the wife of इन्द्रः.

3.3.6 स्व-स्वामिभावः (Owner-property Relation)

स्व-स्वामि relation is marked to indicate owner-property relation. For example - the synsets of वैजयन्तः - the house of इन्द्रः and अमरावती - the city of इन्द्रः are related to the synset इन्द्रः with owner-property relation. इन्द्रः is the स्वामी of वैजयन्तः and अमरावती.

3.3.7 सेव्य-सेवकभावः (Lord-servant Relation)

सेव्य-सेवक relation is marked to indicate lord-servant relation. For example - the synset of गरुडः, the vehicle of विष्णुः is related to the synset विष्णुः with lord-servant relation. विष्णुः is the सेव्यः of गरुडः and गरुडः is the सेवकः of विष्णुः.

3.3.8 धर्म-धर्मिभावः (Property-locus Relation)

धर्म-धर्मि relation is marked in this field. For example - the synsets of वैष्णवी, the power of विष्णुः is related to the synset विष्णुः with property-locus relation. विष्णुः is the धर्मि of वैष्णवी and वैष्णवी is the धर्मः of विष्णुः.

3.3.9 गुण-गुणिभावः (Quality-qualificand Relation)

गुण-गुणि relation is marked in this field. For example - the synsets of श्रीवत्सः, the mark of विष्णुः is related to the synset विष्णुः with quality-qualificand relation. विष्णुः is the गुणि of श्रीवत्सः and श्रीवत्सः is the गुणः of विष्णुः.

3.3.10 उपजीव्य-उपजीवकभावः (Life-livelihood Relation)

उपजीव्य-उपजीवक relation is marked to indicate livelihood. For example - the synset of मत्स्यः is related to the synset धीवरः with life-livelihood relation. मत्स्यः is the उपजीव्यम् of धीवरः and धीवरः is the उपजीवकः of मत्स्यः.

3.3.11 अवतारः (Incarnation)

In this field the incarnation or अवतार relation is marked. For example - the synsets of वामनः, श्रीरामः and श्रीकृष्णः are related to the synset विष्णुः with अवतार relation.

3.3.12 अन्यसम्बन्धाः (Associated With)

This field is meant for other relations which are not defined. For example - the synset देवः may be related to the synset स्वर्गः with a relation आवाससम्बन्धः, is not taken care of. The other relations such as बन्धुता, सौभ्रात्रम्, भ्रातृत्वम् etc. are also not considered here. All such relations are marked as अन्यसम्बन्धाः. These relations will be categorised later.

⁴See Figure 8. in appendix

⁵See Figure 9. in appendix

3.4 Ontological Categories

The ontological categories are handled based on the corresponding ontological charts as described in the जाति: and उपाधि: sections below.

3.4.1 जाति: (Ontological Class)

The universal property of a word is considered as jāti. The ontological categories are marked according to the ontological chart proposed by Nair S S. et. al. (2013). The जाति chart is given in the appendix in Figure 12. Each and every entry has ontological class mentioned in the field⁶.

3.4.2 उपाधि: (Attribute)

Any property ie. qualified to be the universal as per the conditions mentioned in the article of Nair S S. et. al. (2013) is considered as upādhi. The उपाधि classes are marked according to the उपाधि chart proposed. The उपाधि chart is given in the Figure 13 in the appendix.

3.5 Frequency Analysis

For frequency analysis set of 3,000 words are considered. Among them 2876 words are found unique. 2719 words have single sense, 191 words are having two senses, 81 words are having 3 senses and nine words are having four senses. Out of 3000 words 659 Synsets are created. For each word, one or more relations are marked using headwords. Hierarchical relations such as परा-अपरासम्बन्ध: and अवयव-अवयवीसम्बन्ध: are the highly frequent relations. The frequency of high frequent occurrences is detailed in the Table 2.

Relation	Total Words	Total Synsets
परा-अपरासम्बन्ध:	1631	356
अवयव-अवयविभाव:	391	117
जन्य-जनकभाव:	286	15
अन्यसम्बन्धा:	275	83
पति-पत्नीभाव:	175	21
स्व-स्वामिभाव:	149	68
अवतार:	106	15

Table 2: Relational statistics

3.6 Data Implementation

Once the lexicon table was ready with the data, three databases were created using dbm engines of Unix using hashing techniques. Three hash tables were created to represent a data structure to map a given key to value.

- i. Hash table for मुख्यपदम्_headword (key = पदम्_word and value = मुख्यपदम्_headword)
- ii. Hash table for synset (key = मुख्यपदम्_headword and value = synset)
- iii. Hash table for पदम्_word info (key = word and value = निगम:_Reference & लिङ्गम्_Gender)

With the help of this data structure, a user can key in a desired word and get output in the form of synonyms, meaning and related information about the word.

The ontological structure adopted for creating the web-tool for Amarakośa has been replicated here with modifications for two reasons. Firstly, the division of various kāṇḍas and categorization of words in both the lexicons are very similar and therefore what has been

⁶See Figure 6. in appendix

created earlier can be easily adapted to VK as well. Secondly, this also enables for technical integration of the two tools in future that will facilitate easy cross-reference.

As this is an ongoing project, the other relations will be supplied in due course as appropriate. It was felt that the first step to get the synsets in order will provide vital wealth of information to researchers and students on this lexicon and the emphasis was thus on categorizing the headword first.

3.6.1 Processing Flow

The input word, the type of requested information (meaning and relation with other words) along with parameters like input encoding for the identification of input and output encoding for formatting of the result on the webpage is processed by a series of scripts in the server. The scripts identify the word and the relation for which the information is requested. They access the databases corresponding to the relations that have been created a-priori, and extract information from the database(s) corresponding the selected relation and format the output into HTML file.

When “All Relations” is chosen as the input, a pictographic representation of all relations is created and embedded in the resultant HTML file. Refer Figure 11. for the output in the appendix.

This HTML file is returned as response to the requesting browser for display to the user. The following flowchart describes the steps in the processing in Figure 3.

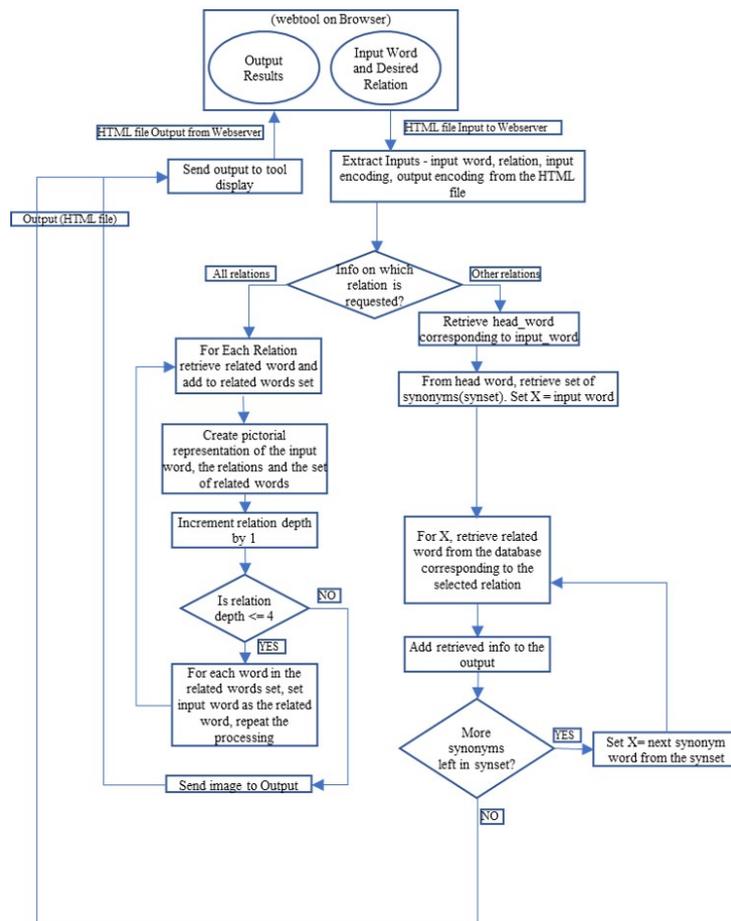


Figure 3: Flowchart for processing in the web tool

3.7 Architecture of VKN

Figure 4 illustrates the architecture of the Vaijayantīkośa KnowledgeNet (VKN) tool. It consists of the following functional components.

- Web user interface
- Webserver
- VK datasets



Figure 4: Architecture of the VKN tool

3.7.1 Web User Interface

The user interface for the tool is a HTML web page (currently, first version of the tool is available at <http://13.235.131.68/CompLing/vk/>) It provides a means to input the word from the VK lexicon that needs to be analysed for the specific set of relations. See Figure 5.

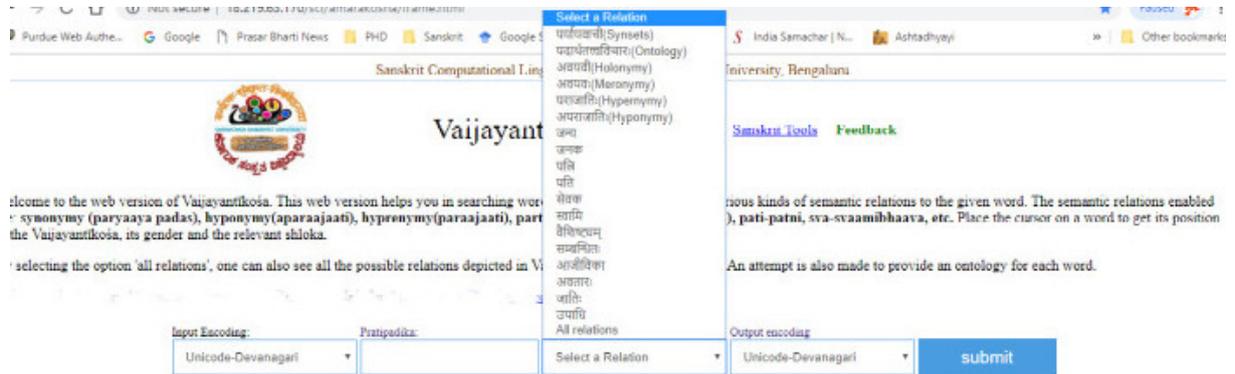


Figure 5: VKN tool

Multiple input encoding forms are provided including Devanagari and WX encoding. The input word can either be प्रातिपदिकम् or प्रथमा एकवचनरूपम्. The desired semantic relation can be extracted from the lexicon from the drop-down list. The tool supports analysis of the relations mentioned in the section 3.3.

3.7.2 Webserver

An Apache webserver hosted on an ubuntu instance running on AWS, is used to interact with the Web user interface. It captures the inputs from the HTML webpage and passes onto the CGI script in the backend for processing. The result of the processing is sent as HTML response to the requesting webpage for display to the user.

3.7.3 VK Datasets

WX encoded original VK ślokas and database (see section 3.2 and 3.3) that contain manually created and verified metadata for each word are the input files. The databases are created as per the data implementation described in section "Data Implementation". The processing scripts analyse the inputs for requested information/relations associated with the words in VK, retrieve desired information from these datasets and display the results in the tool as results.

3.8 VKN Android Application

VKN Android App provides a convenient interface to Android smartphone users to access and analyse information in the Vyjayantīkośa. It uses the same input data set, words and relation information used for the web-tool. The Android App collects inputs from the user i.e - the word and its relation. It then communicates the input parameters to a webserver hosted in the cloud, where python scripts are used to search and formulate the response using the input data set. The response is conveyed back to the App on the smartphone for display.

The VKN android App is available for download from the VKN tool webpage. The tool is under development and has been released for volunteer testing and collecting feedback on usability. It currently allows input in Devanagari format and supports the synset relation analysis. The App is being enhanced to support relations and features supported by the web tool as discussed in the previous sections.(See Figure 10)

4 Conclusions

VK has a rich repository of words from the Sanskrit language and literature. The VKN web-tool enables convenient access to this knowledge. It is also designed to enable analysis in specific areas of research by providing a list of words related to that area, which can be used to trace information related to that area in Sanskrit literature. For example, in a paper published in the Indian Journal of History of Science, the use of the term hemaghna (destroyer of Gold), for lead metal was examined in detail (Dube, 2010), and this uncovered, unique properties of the metal lead, when interacting with Gold. There is scope for deeper research for experts from different fields - geology, geography, ornithology, metallurgy, sociology, biology and more. In this context, this tool becomes significant as it provides preliminary information to researchers in their respective fields with synsets and ontological structure and could become a starting point for a more comprehensive research. The inclusion of meaning in English, bridges the language divide, connecting this knowledge base with the large number of English speaking researchers.

5 Future Research

Few suggested future work is as follows:

- Continue updating the kośa with all the remaining entries.
- The child-parent, master-possession, husband-wife relations and other such relations (see section 3.2) were captured at this stage. There are possibilities of including other relations such as siblings, dwellings etc.
- Linking each synset to Amarakośa Knowledge Net.
- Linking it with various other computational linguistic tools.
- Using for Word sense disambiguation.
- Currently only four layers of nesting depth is represented in "all relations". This can be expanded to more layers in future.

Acknowledgement

The Authors would like to thank Prof. Amba Kulkarni, Department of Sanskrit Studies, University of Hyderabad, Hyderabad for providing technical support at various levels. The project team also acknowledges the valuable guidance of Prof. Shrinivasa Varakhedi, Vice-chancellor, Kavikulaguru Kalidasa Samskrita University, Nagpur. The team of Post Graduate Diploma in Sanskrit Computational Linguistics at Karnataka Samskrit University, Bangalore is also acknowledged.

References

- Bühler, G. (1887). Gleanings from Yâdavaprakâsa's Vaijayantî. Wiener Zeitschrift für die Kunde des Morgenlandes, 1, 1-7. Retrieved April 1, 2019, from <https://www.jstor.org/stable/23858800>
- Dube, R. K. (2010). An Assessment of the Sanskrit Word Hemaghna used for Lead Metal. Indian Journal of History of Science, 395-401.
- Haragovindashastri, P. (1971). Vaijayantikośa. Varanasi: Chowkhamba Sanskrit Series Office.
- Kaur, S., & Singh, L. (2018). Indian Arthropods in Early Sanskrit Literature: A Taxonomical Analysis. Indian Journal of History of Science, 59-64.
- Kulkarni Amba. & Nair S Sivaja. (2010) Knowledge Structure in Sanskrit Kosas. Proceedings of 8th ICON, Indo-wordnet Workshop, IIT Khragpur.
- Nair, S Sivaja. (2011). The Knowledge Structure in Amarakośa. Hyderabad: University of Hyderabad.
- Nair, S Sivaja, Varakhedi Shrinivasa & Shivani V (2013). Extended Nyaya-Vaisesika Ontology as Applied to Amarakośa KnowledgeNet, Proceedings of 5th ISCLS, IIT Bombay, DK Print world, New Delhi
- Oppert, Gustav. (1893). The Vaijayanti of Yadavaprakasa. Madras: Madras Sanskrit and Vernacular Text Publication Society.
- Patkar, M. M. (1981). History of Sanskrit Lexicography. New Delhi: Munshiram Manoharlal.
- Popescu, F. (2019). A Paradigm of Comparative Lexicology. UKNewcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Sun, M. H., Safwanah, N. L., & Tan, D. (2017). Lexicology: The Importance of Words in Society. Universiti Sains Malaysia. Retrieved April 19, 2019, from https://www.researchgate.net/profile/Ernest_Mah/publication/320839664_Lexicology_The_Importance_of_Words_in_Society/links/59fcb88baca272347a22773b/Lexicology-The-Importance-of-Words-in-Society.pdf
- Varakhedi, Shrinivasa., Jaddipal, Viroopaksha. & Sheeba, V. (2007). An effort to develop a tagged lexical resource for Sanskrit. Sanskrit Computational Linguistics, 339-345. Retrieved April 19, 2019, from <https://hal.inria.fr/inria-00207962/d>

A Appendix - 1

VKN Sample Outputs

अर्थः :: पक्षी | वर्गः :: खगाध्यायः | , पक्षिन्, पत्ररथ, पत्रिन्, पित्सत्, पिपतिषत्, पतत्, पतङ्ग, पतग, प्लाविन्, पतत्रिन्, अङ्गु, पतत्रि, विहङ्गम, विहग, विहङ्ग, नभसङ्गम, नीडोद्भव, शुक, नीडिन्, मलूक, विप्रुष, भसत्, वशाकु, मदन, पीतु, मशाक, मदुर, द्विज, ऊक, शकुन्ति, शकुनि, शकुन्त, शकुन, खग, शलक, विकिर, तुण्डिन्, नीडज, वातगामिन्

जाति
=> पक्षी
=> जन्तुः
=> मनुष्येतरः
=> चलसजीवः
=> पृथ्वी
=> द्रव्यम्
=> पदार्थः

Figure 6: Example of Ontology

अर्थः :: अग्निः | वर्गः :: लोकपालाध्यायः | , वह्नि, वैश्वानर, घासि, कृष्णवर्त्मन्, समन्तभुज, जातवेदस्, बृहद्राणु, वीतिहोत्र, तनूनपात्, दहन, ज्वलन, शुष्मन्, रोहिताश्व, उषर्बुध, शोचिष्केश, त्रिधामन्, अग्नि, उदर्विस्, पावक, अनल, हिरण्यरेतस्, सप्तार्चिस्, वसुरेतस्, हुताशन, कृपीटयोनि, अर्विष्मत्, धूमकेतु, दुरासद, मन्त्रजिह्व, सप्तजिह्व, सुगिह्व, हव्यवाहन, आश्रयाश, वातसख, कृशानु, वातसारथि, वभि, भुजि, पचि, साधि, विरि, वसति, अश्वति, जागृवि, सहुरि, सदिधे, दनुमस्, हवन, हव, सुवाकु, भरथ, पीथ, जुहुराण, ईषिर, आशिर, तेजस्, अप्पित्त, अपांपित्त

अवयवः (Meronym)

अर्थः :: धूमः | वर्गः :: लोकपालाध्यायः | , धूम, तरि, मेघवाहिन्

अर्थः :: जिह्वा | वर्गः :: लोकपालाध्यायः | , शिखा, जिह्वा, कील, ज्वाल

अर्थः :: स्फुलिङ्गः | वर्गः :: लोकपालाध्यायः | , स्फुलिङ्ग, अपुञ्ज, खड्गाङ्ग

अर्थः :: सन्तापः | वर्गः :: लोकपालाध्यायः | , सञ्चर, सन्ताप

अर्थः :: उल्का | वर्गः :: लोकपालाध्यायः | , उल्का, क्रमुक, अलात, उत्सुक

अर्थः :: भस्मन् | वर्गः :: लोकपालाध्यायः | , भस्ति, भस्मन्, भूति

Figure 7: Example of अवयवः

अर्थः :: वायुः | वर्गः :: लोकपालाध्यायः | , वात, वायु, जगत्प्राण, शुषिल, श्वसन, अनिल, गन्धवाह, गन्धवह, मातरिश्वन, समीरण, युजिन, लघुग, वाति, ध्वजप्रहरण, जगत्, दैत्यदेव, वह, प्राण, नभःप्राण, अङ्कति, सर, आशुग, पवन, स्पर्श, पवमान, प्रभञ्ज, मरुत्, धूलिध्वज, मर्क, समीर, अग्निसख, चल, शुष्मि, महाबल, वेगिन्, नभोजात, सदागति, नभस्वत्, मारुत्, शीघ्र, पृषदश्च, प्रकम्पन

अपराजाति (Hyponym)

अर्थः :: सङ्क्रा | वर्गः :: लोकपालाध्यायः | , सङ्क्रा

अर्थः :: वात्या | वर्गः :: लोकपालाध्यायः | , वात्या

अर्थः :: मृदुवातः | वर्गः :: लोकपालाध्यायः | , मृदुवात, चिञ्चिलिक, लेढ, लिह, सौरत

अर्थः :: हारितः | वर्गः :: लोकपालाध्यायः | , हारित, नेरिन्, यौनिक, स्वेदचूषक

अर्थः :: झञ्झानिलः | वर्गः :: लोकपालाध्यायः | , स्वरिङ्गण, झञ्झानिल

Figure 8: Example of अपराजातिः

अर्थः :: कार्तिकेयः | वर्गः :: आदिदेवाध्यायः | , कुमार, शरज, स्कन्द, तारकारि, उमासुत, महासेन, महातेजस, शक्तिपाणि, गुह, अग्निभू, वैजयन्त, सिद्धसेन, सेनानी, शिखिवाहन, स्वामिन्, गाङ्गेय, गौरेय, ब्रह्मचारिन्, महौजस, षाण्मातुर, कार्तिकेय, ब्रह्मगर्भ, षडानन, सुब्रह्मण्य, नीलवट्ट, क्रौञ्चारि, कुक्कुटध्वज

जनक

अर्थः :: शिवः | वर्गः :: आदिदेवाध्यायः | , महेश्वर, पशुपति, श्रीकण्ठ, पांसुचन्दन, शङ्कर, गिरिश, रुद्र, गिरीश, शशिभूषण, भद्रेश्वर, चन्द्रमौलि, पिनाकिन्, शशिशेखर, कपर्दिन्, धूर्जटि, शर्व, कपालिन्, नीललोहित, ईश, ईश्वर, ईशान, भर्ग, मृत्युञ्जय, मूड, व्योमकेश, महादेव, प्रमथाधिपति, शिव, शूलिन्, दक्षाध्वराराति, कामारि, परमेश्वर, कृत्तिवासस, अहिर्बुध्न्य, नीलग्रीव, त्रिलोचन, गङ्गाधर, विरूपाक्ष, वामदेव, वृषध्वज, भूतेश, खण्डपरशु, स्थाणु, अन्धकसूदन, भगनेत्रान्तक, भीम, त्रिपुरारि, दृगायुध, कटाटङ्क, जटाटीर, जटाझाट, महानट, झिण्टीकान्त, रेरिहाण, सर्वज्ञ, नन्दिवर्धन, स्थाल, डिण्डीश, उड्डीश, कण्ठकाल, महाव्रत, कृशानुरेतस, जोटिङ्ग, कङ्कटीक, उमापति, खड्गङ्गिन्, मन्दरमणि, उलन्द, वृषवाहन, उग्र, कटप्र, दिवासस, झण्ड, षण्ड, अकृतश्चन, बहुरूप, अर्धमुकुट, दशबाह, दशाव्यय

Figure 9: Example of जनकः

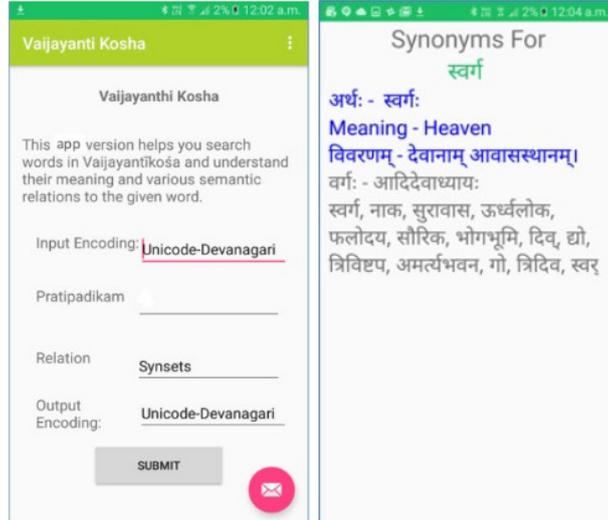


Figure 10: VKN Android Application

उपाधिवृक्षः

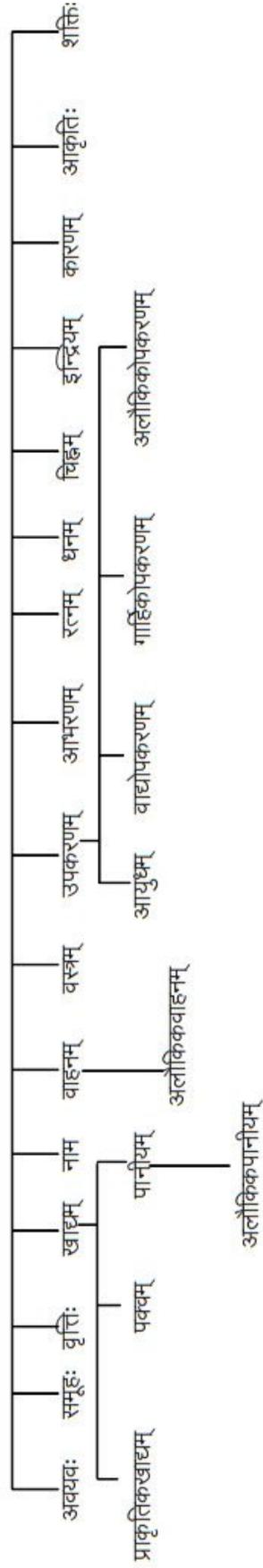


Figure 13: Upādhi Chart