

# A Web-Based Sanskrit Verb Thesaurus

## BTP Review 1

Aditya Srinivas Menon  
Aditya Garg  
Aditya Tyagi  
Tarun B

Indian Institute of Information Technology Kottayam

Guided by: Dr. Nandini J Warriar

13th September 2023



# Outline

- 1 Introduction
- 2 Motivation
- 3 Problem Statement
- 4 Literature Survey
- 5 Proposed Approach
- 6 Conclusion



# Introduction

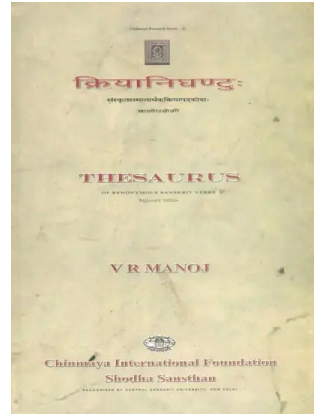
## Revolutionizing Sanskrit Language Understanding

### Sanskrit

- "Sanskrit as a Language is Complex" [1]
- All languages in the Indo European family in India, the roots can be traced to Sanskrit
- Sheer volume of lexical knowledge accumulated over at least 2000 years
- Intricate grammar and vast vocabulary
- Culturally strong language rich with heritage
- Tons of available data in the form of scriptures

# Kriyānighantu: The Sanskrit Verb Thesaurus [2]

- Compiled and published by Manoj VR
- Thesaurus of Sanskrit verbs, to help Sanskrit enthusiasts find appropriate verbs and their meanings
- Many sanskrit words acquire new meaning over time
- Having a Verb Thesaurus is crucial for grammarians to understand the text how it was written
- 664 pages of Verbs with forward and backward links
- The author is praised for his dedication to Sanskrit research



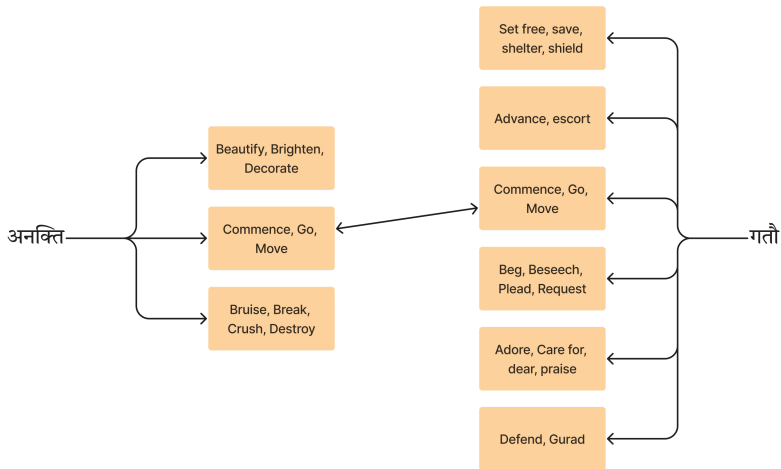
# Problem Statement

- Although there exists a Sanskrit wordnet [1], it contains Nouns, Adverbs and Adjectives and give less importance to the rich Verbology of Sanskrit
- Existing Sanskrit wordnet has limited number of words
- To solve this problem we leverage the book to build a online, web-based version Sanskrit verb thesaurus
- Use this web based verb thesaurus to make Sanskrit language more accessible

# How complex is the task?

- Sanskrit verbology is vast system with over 1993 root words
- These root words can be combined to make a large number of combinations
- A challenging task for any scholar to go understand grammar of Sanskrit text

# How complex is the task?



# Advantages of our approach

- **Web-Based is Better:** Accessibility and convenience, no need for physical books.
- **Easier Searching:** Quick access to specific verbs.
- **Cost Efficiency:** No expenses on physical resources.
- **Simplifying Sanskrit Grammar:** A valuable tool for Sanskrit grammarians.
- **Preservation of Cultural Heritage:** Facilitating the understanding of ancient texts.



# Existing Approaches

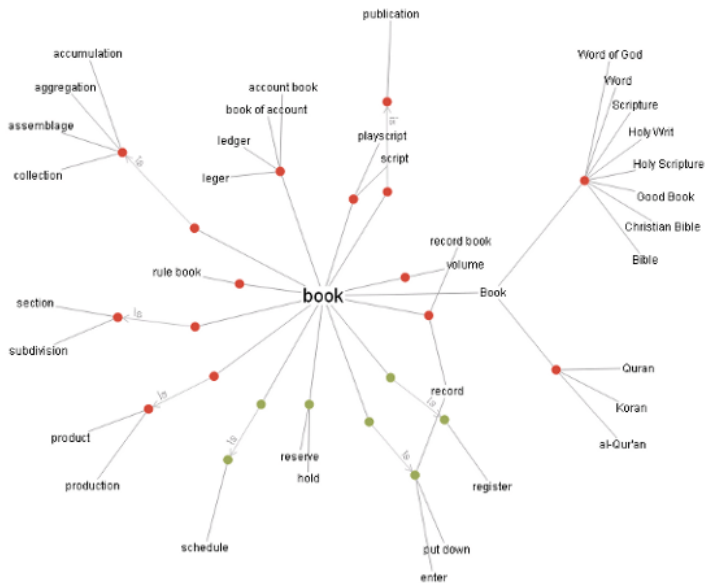
## What is a Wordnet?

A WordNet is a lexical database that organizes words based on their meanings and semantic relationships created by Princeton University

There are 3 primary components of Wordnet

- **Synsets (Synonym Sets):** The core building blocks of a WordNet are synsets, which are groups of words or phrases that are synonymous or have closely related meanings.
- **Lexical Relations:** WordNet encodes various lexical relations that describe the connections between synsets

## Wordnet Example



# Advantages of Wordnet

- **Word Sense Disambiguation (WSD):** Wordnet gives us an inventory of words and their senses, making it easier to disambiguate the meanings of words in context.
- **Semantic Similarity and Relatedness:** WordNet encodes semantic relationships between words, such as synonyms, antonyms, hyponyms (more specific terms), and hypernyms (more general terms).
- **Information Retrieval:** WordNet can improve information retrieval systems by expanding query terms to include synonyms and related words
- **Machine Translation:** Having access to related words can be useful in translation
- Language Understanding
- General resource for NLP research

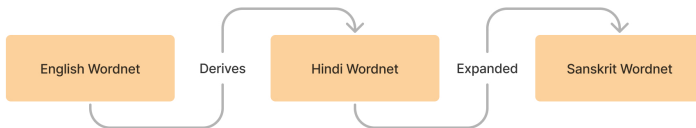
# Sanskrit Wordnet: Pushpak Bhattacharya [1]

- Hindi Wordnet was expanded to form the Sanskrit wordnet. Hindi wordnet was initially made from the English Wordnet
- The hierarchy preservation principle (HPP) is followed in the Sanskrit Wordnet, which means that the hierarchy of the Hindi Wordnet is used as a reference to establish semantic relations for the synsets of Sanskrit Wordnet
- In case of nouns all gender variations are included in the synset

# Sanskrit Wordnet: Pushpak Bhattacharya [1]

Distribution of POS tags

Noun	Verb	Adverb	Adjective
17413	1246	263	3990



# Sanskrit Wordnet: Pushpak Bhattacharya [1]

## Disadvantages

- Only 1246 verbs, Sanskrit is a language with a rich Verbology
- Room for improvement of User Interface.
- Aging technology and research. Uses old database schema.
- Detailed explanations of the methods and techniques used in building the Sanskrit Wordnet is not mentioned
- The paper does not discuss about the evaluation of the Sanskrit Verbnet
- Does not provide information on the practical applications of the Sanskrit Wordnet, such as its use in natural language processing tasks.

# JNU tanalyser [3]

- Morphological analysis methodology to develop a computational system that can identify and analyze Sanskrit verb forms.
- The backend of the Sanskrit verb analyzer system contains lexical resources in the form of data-files.
- Follows a multi-tier architecture that involves the user, Apache-tomcat, Java servlet, and data files.



## Disadvantages

- Architecture based on old tools and frameworks
- Does not work most of the time
- Morphological analysis leads to very little verbs identified. We have a book that deals with a lot of verbs and acts as a rich database



# Other approaches

## Other approaches

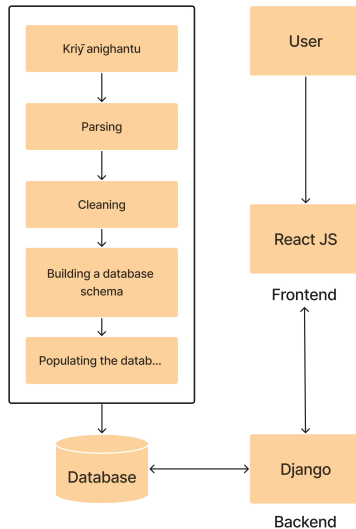
- Many of the approaches from 2000-2010 that try to build Sanskrit wordnet have been discontinued
- The website does not open or approaches have not been mentioned
- [4] and [5] are some of the examples
- Sanskrit Grammarians cannot find a working online thesaurus for verbs

# Proposed Approach

- Step 1: **Digitalization:**  
Convert the book into a digital format.
- Step 2: **Database Schema**  
Create a structured database schema.
- Step 3: **Data Conversion**  
Populate the database with book content.
- Step 4: **User-Friendly Interface**  
Develop a frontend for intuitive verb retrieval.
- Step 5: **Document Search**  
Input a Sanskrit document, and our app will annotate all verbs.



# Framework



# Conclusion

Revolutionizing Sanskrit Language Understanding by making an online version of the Verb thesaurus written by Manoj VR

- Transforming complex Sanskrit into an accessible, digital resource.
- Bridging the gap between tradition and technology.

# References

- [1] M. Kulkarni, C. Dangarikar, I. Kulkarni, A. Nanda, and P. Bhattacharyya, "Introducing sanskrit wordnet,".
- [2] V. Manoj, - *Thesaurus of Synonymous Sanskrit Verbs*. Chinmaya International Foundation Shodha Sansthan, 2021.
- [3] [Online]. Available:  
<http://sanskrit.jnu.ac.in/tanalyzer/tanalyze.jsp>.
- [4] [Online]. Available:  
<https://sanskrit.inria.fr/DICO/grammar.html>.
- [5] [Online]. Available: <http://sanskrit.uohyd.ernet.in/>.