Problem definition and scope of work

Literature survey

Proposed architecture

Applications of proposed work

Presentation skills

**Slide 1: Introduction**

* Title: “A Web-Based Sanskrit Verb Thesaurus”
* Subtitle: Revolutionizing Sanskrit Language Understanding

**Slide 2: Problem Statement: Introduction**

* “Sanskrit as a Language is Complex”
* 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

**Slide 3: The Book**

* “Kriyānighantu” by VR Manoj.

**Slide 4: Problem Statement**

* We aim to create an online, web-based version of this Sanskrit verb thesaurus.

**Slide 5: The Complexity of the project**

* “How Complicated is it?”
  + A vast system: Over 1993 root words.
  + Huge number of verb combinations.
  + A challenging task for any scholar

**Slide 6: The complexity of the project**

* (example)

**Slide 7: Advantages of Our Approach**

* “Advantages”
  + “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.

**Slide 8: Existing Approaches**

* “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.
    - Advantages of Wordnet in NLP
      * 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. (Example)
      * 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
    - How is a wordnet created

**Slide 9: 1. Sanskrit Wordnet Pushpak Bhattacharya**

* “Sanskrit WordNet”
  + <https://www.cfilt.iitb.ac.in/wordnet/webswn/english_version.php>
  + 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
  + Distribution of POS tags are as follows

| Noun | Verb | Adverb | Adjective |
| --- | --- | --- | --- |
| 17413 | 1246 | 263 | 3990 |

English Wordnet → Hindi Wordnet → Sanskrit Wordnet

**Slide 10: 1. Sanskrit Wordnet Pushpak Bhattacharya**

* 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.

**Slide 11: JNU tanalyser**

* Advantages
  + 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.

**Slide 12: JNU tanalyser**

* 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

**Slide 13: 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
* <https://sanskrit.inria.fr/DICO/grammar.html>
* <http://sanskrit.uohyd.ernet.in/>
* Sanskrit Grammarians find it hard to choose a thesaurus for Verbs

**Slide 14: Our Approach**

* “Our 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.

**Slide 15: Conclusion**

* “Revolutionizing Sanskrit Language Understanding”
  + Transforming complex Sanskrit into an accessible, digital resource.
  + Bridging the gap between tradition and technology.

## PPT

1. Define problem statement
   1. Sanskrit as a language is complex
   2. Book: Kriyānighantu is a sanskrit VERB thesaurus
   3. Author: VR Manoj
   4. Online web based version of Sanskrit verb thesaurus
2. How complicated it is
   1. Add verb example, 1993 root words and no of verb combinations cannot be counted
3. Advantages
   * 1. Web based is better than book
     2. Searching is easier
     3. Don't have to spend money on books. As it is cost efficient
     4. Life of a Sanskrit grammarian is easier with this tool
     5. Cultural heritage preservation by understanding old texts
4. Existing Approaches
   1. What is a wordnet?
   2. Sanskrit Wordnet
      1. Nouns too, What we are building is a verbnet
      2. Interface can be better
      3. Old
5. Approach
   * 1. Get book in digital format
     2. Build a schema for database
     3. Convert book to database based on the schema
     4. Build a frontend for graph based node retrieval and display of the verbs
     5. Document search: Input a sanskrit document and the app will annotate all the verbs