



Ontological Knowledge base for Selected Verbs of Sanskrit and Bangla

Subhash Chandra

subhash.chandra@cdackolkata.in, subhash.jnu@gmail.com

Special Centre for Sanskrit Studies, Jawaharlal Nehru University (JNU), New Delhi-110067, India

1. Introduction

- To develop an ontological knowledge base for selected verb of Sanskrit and Bangla
- Study of relationship between verbs and their acceptable arguments
- Exploring theories of how verb semantics can determine the morpho-syntactic valence of its arguments
- The work constructs a bridge between first argument (subject) and second or other argument (object) for semantic and syntactic research
- This paper presents a method to represent knowledge in machine for identification of valid subject and object in a Sanskrit sentence

4. Expectation principle of Verb and Compatibility

Expectation of Verb

Intransitive verbs expect one noun as an argument (subject), and transitive verbs expect more than one noun as arguments with valid compatibility between subject and object/s

Compatibility of Verb Arguments

Verb expects compatible subject/object/s for action to be complete

8. Technical Details

- Web based data entry interface has been developed in Python.
- MySQL database and text files for the backend
- Python Server Pages (PSP) for the front-end
- Python for the programs

2. Problem

- To understand verb argument relationship by machine, we need to represent real world knowledge in the machine.
- Human brain knows a verb has certain expectancy for its arguments whereas noun has a mutual compatibility for associating with a specific verb.
- Machine don't have such type knowledge

5. Databases

Knowledge Database of Sanskrit and Bangla

It is knowledge database with property information of each base word (right now nouns and pronouns only) collected from corpora and MWD for Sanskrit

Argument Valence Mapper Database

- It is subject and object mapper database
- All acceptable combinations of argument for the specific verb are given in the this database

9. Analysis of Verb

- This system is under development at this point.
- The analysis is done with the help of relational databases.
- The system gives the ontological information of each input word with the help of relational database.
- Checks all ontological information in the Mapper.
- If any combination of arguments is match then the system returns it as valid combination of argument, otherwise it labeled invalid.

3. Sanskrit Verb

- About 2000 verb roots classified in 10 morpho-semantic classes called *gaNas* (group)
- These can have two major semantic classes called *aatmane* and *parasmai* based on who the beneficiary of the action is and also whether something is a universal fact
- Can also be potentially prefixed with 22 prefixes

6. User Interface

Data Entry Interface

- The data entry user interface is web based.
- The interface allows users to modify existing entry and enters new words with ontological information in database with the help of ontological tree viewer
- The first text area is for searching whether the entry exists or not, the second is for updating or inserting in the database and the tree viewer is for ontological information entry.

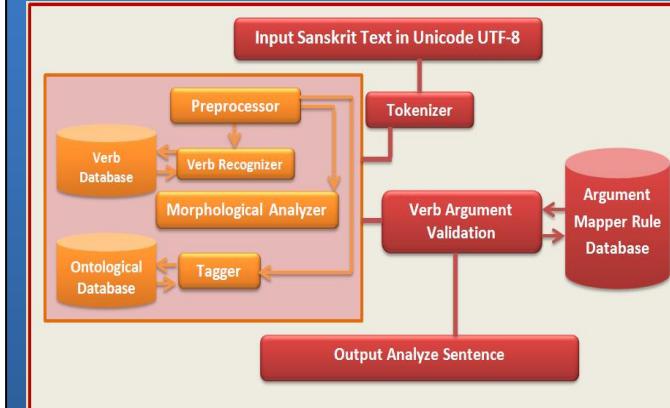
Analysis of Verb Arguments Valence System Interface

- This interface accepts Sanskrit sentence, text or corpora in utf-8.
- User can also upload a Unicode text file through file browse.
- After clicking on the button, the system returns analyzed data with acceptable subject and object tags

7. Methodology

- Developing the knowledge database using the ontological method
- Developed a ontological tree with the help of available tree structure, self-knowledge and corpora
- The tree is based on Sanskrit verb expectancy (subject and object structure)

Overall System can be understand through following flow chart-



10. Conclusion

- This is an effort to providing knowledge to the machine for verb argument valency.
- The current work is in developmental stage at this point, so it does not cover all Sanskrit verbs.
- Though, this approach does not promise a complete solution, yet it may be a very effective model for language processing in general.
- It will be a model for Indian languages for computational processing and can be very useful system for knowledge representation in machines which has been very challenging till date.