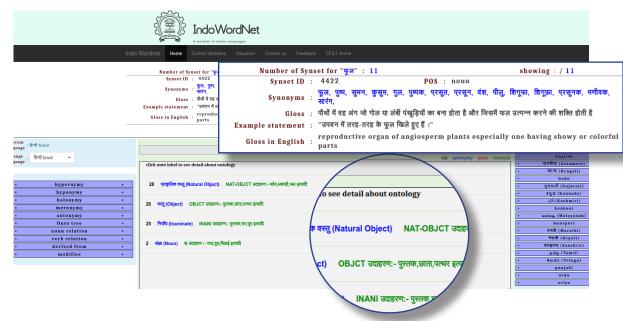# IndoWordNet

WordNets are lexical structures composed of sets of synonyms called synsets, and semantic relations between these synsets. Wordnets help in various natural language processing (NLP) tasks such as word sense disambiguation, machine translation, etc. Unavailability of a crucial lexical resource like Wordnet has impeded the development of NLP technologies for Indian languages.

Funded by several key government agencies like Technology Development for Indian Languages, Ministry of Electronics & Information Technology, Ministry of Communications & Information Technology and Ministry of Human Resource Development, the Centre for Indian Language Technology (CFILT) at IIT Bombay has been a flag-bearer of the IndoWordnet project. IndoWordNet is a linked structure of Wordnets of major Indian languages from Indo-Aryan, Dravidian and Sino-Tibetan families. These Wordnets have been created by following the expansion approach from Hindi wordnet which was made available free for research in 2006. Since then, a number of Indian languages have been creating their Wordnets. These Wordnets have been used extensively for NLP in Indian Languages.

CFILT has been leading this effort pan India with the help of a consortium, where universities across the country have participated in building the structure. IndoWordNet consists of 19 Indian Languages including English, and boasts of approximately 40000 synsets where more than 20000 have been linked for all the languages. WordNet data has been made available for download for the purpose of research.

A current branch of work that CFILT is exploring deals with use of Wordnet for computational phylogenetics for Sanskrit, where manuscripts can be analysed textually for creating a critical edition of the work. It can be used to study word etymology, morphological structure of the word, and predict properties of the manuscript such as date of creation, author attribution, and when a particular version of the manuscript was created. This helps in placing ancient manuscripts over a timeline, and reaching their origins.

Prof. Pushpak Bhattacharyya, Department of Computer Science and Engineering, pb@cse.iitb.ac.in