Abstract

India is a country with diverse languages. Many a times, words are spoken and used which have more than one meaning. While interaction between humans, the determination of the correct meaning of the ambiguous word (word with multiple meaning) can easily be judged by referring the context of the communication. But for a computer to judge the best and the correct meaning of the word, training needs to be provided to the system. Word Sense Disambiguation (WSD) is technique used to disambiguate the ambiguous words (single word with multiple meaning). Our work deals with analyzing the correct meaning of the ambiguous word(s) in Punjabi language. Not much work has been done in this field which deals with the Punjabi language. A text with multiple senses in natural language is open problems of Natural Language Processing (NLP) which can be resolving using WSD. The Supervised learning methodology is used for this purpose which is the conventional approaches to WSD. The semantic lexicon for the various languages of India i.e., Indo WordNet has been used to obtain the sense definition of the Punjabi language. An enhanced Lesk approach is used to analyze the correct sense of the ambiguous word which uses the concept of dynamic context window. The proposed
algorithm works on one assumption that the words on the left and the right of the context window should be of the same theme in its neighborhood. Finally, instance and precision is obtained which shows that larger the size of context window, more appropriately the correct sense of the word can be determined.

References


Index Terms

Computer Science  Pattern Recognition

Keywords

Word sense disambiguation, Lesk, Punjabi Language, Natural language processing