Abstract[1]

Words are fundamental in semantic analysis of natural language expressions. An important challenge in the task of semantic analysis is the fact that in natural languages several words may denote the same concept, and a single word form may denote different concepts. It is thus important to have information about such characteristics of words, in a readily accessible format. WordNet is such a resource that may be developed for a language. It is a database of words in a language, arranged in the form of sets-of-synonyms (synsets). Various types of lexical and semantic connections between different synsets are also recorded in the WordNet. WordNet was originally developed for the English language. Very soon WordNets for other languages of the world also were developed. In India too, WordNets for many official languages are being developed; some of these are in a very initial state. Ours is one of the first such efforts for the Assamese language, a language spoken by nearly 30 million people. Besides WordNet for specific languages, multilingual WordNets covering several languages are also being developed, such as the EuroWordNet. In this line, in India an IndoWordNet is being conceptualised. In order to develop a complete multilingual IndoWordNet, there should be a WordNet for each Indian language, but till date, there is no WordNet for Assamese language. Our work mainly focuses on developing the WordNet for Assamese language. An exercise such as this calls for efforts in the domain of linguistics as well as computer science. The WordNet database consists of text files which include the synonym set of a word and the sense of the word in multiple languages. Synsets are interconnected with other synsets via a number of lexical and semantic relations. In this dissertation, proposed file formats of the Assamese WordNet database text files are discussed. We also describe about the Application Programming Interface (API) of the Assamese WordNet through which other applications can access the database. Up to the time of preparation of this dissertation, about 2000 words in more than 300 synsets are present in the Assamese WordNet.