Word Sense-based Information Retrieval (WSIR) by using Jaccard Coefficient Similarity Method

Thaw Zin Myint, Nilar Aye
University of Computer Studies, Yangon
thawzinmyint.lmn@gmail.com

Abstract

Nowadays, information retrieval (IR) system is to provide users with documents that contain their information need based on the user query. Due to the ambiguous query words, the keyword-based IR system fails to recognize the relevance of documents to the query. Word sense disambiguation (WSD) can solve this problem to improve the keyword-based IR performance. So, this system proposes the word sense-based IR (WSIR) system by using Jaccard coefficient based similarity method. In this system, similarity based WSD method is used to disambiguate the ambiguous words in the user query. This system also uses the WordNet and Corpus as the lexical resources that encoded senses of each word. In the WSIR system, the various senses that are provided by the WSD method have been used as semantics for indexing the documents. To show the better performance of the WSIR, this system compares the keyword-based IR and word sense-based IR. This system is implemented by using C# programming language.