Abstract

Text clustering plays an important role in providing intuitive navigation and browsing mechanisms by organizing large sets of documents into a small number of meaningful clusters. Many fuzzy clustering algorithms, such as K-means, deal with documents as bag of words. The bag of words representation method used for these clustering is often unsatisfactory because it ignores the semantic of words. The proposed agent exploits WordNet ontology to create low dimensional feature vector which allows us to develop an efficient clustering algorithm. A new semantic-based model, that represents documents based on semantic concepts of words, is proposed. The proposed approach aims at increasing the performance of information retrieval process by enhancing the document clustering. The accuracy and the speed of clustering have been examined before and after combining ontology with Vector Space Model (VSM). Experimental results demonstrate that using semantic-based model and fuzzy clustering enhances the clustering quality of sets of documents.

References


Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (Dublin Ireland, 1994), 61. ACM.

Index Terms

Computer Science
Fuzzy Systems
Keywords
   Document Clustering  Semantic Text Representation  Agent  WordNet