Abstract

Document clustering is an automatic grouping of text documents into clusters so that documents within a cluster have high similarity values among one another, but dissimilar to documents in other clusters. It has wide applications in areas such as search engines, web mining, information retrieval and topological analysis. This paper presents a new document clustering algorithm using the concept of fuzzy sets, where each cluster is viewed as a fuzzy set over some finite universal set. The algorithm was implemented and the results are reported. The efficiency and time complexity of the algorithm have also been discussed.

References

3. J. C. Bezdek, “Pattern Recognition with Fuzzy Objective Function Algorithm”, Plenum


Index Terms

Computer Science        Pattern Recognition

Keywords

Document Clustering, Fuzzy Set, Agglomerative Algorithm, Compact Representation.