Abstract

Document clustering is one of the difficult and recent research fields in the search engine research. Most of the existing documents clustering techniques use a group of keywords from each document to cluster the documents. Document clustering arises from information retrieval domains, and "It finds grouping for a set of documents belonging to the same cluster are similar and documents belongs to the different cluster are dissimilar." The information retrieval plays an important role in data mining for extracting the relevant information for related to user request. Information retrieval finds the file contents and identifies their similarity. It measures the performance of the documents by using the precision and recall. In this paper we proposed a phrase based clustering scheme which based on application of Suffix Tree Document Clustering (STDC) model. The proposed algorithm is designed to use the STDC model for accurate equivalent representation of document and similarity measurement of the similar documents. This method of clustering reduces the grouping time and similarity accuracy as compared to other existing methods.

References
Phrase based Clustering Scheme of Suffix Tree Document Clustering Model

- Shafiq Alam, Gillian Dobbie, Patricia Riddle, M. Asif Naeem, "Particle Swarm Optimization Based Hierarchical Agglomerative Clustering", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 64-68.

**Index Terms**

Computer Science

Artificial Intelligence

**Keywords**

Clustering Techniques  Document Clustering  Phrase Merging  Suffix Tree