Abstract

A rapid growth of documents available on the Internet, digital libraries, medical documents, news wires and other scientific document corpuses has motivated the researchers to propose many text mining techniques that help users to quickly retrieve trace and summarize the information in an effective way. Topic detection is one such technique which discovers precise, meaningful and concise labels for the formulated static document clusters. This technique helps the user to navigate and retrieve the needed information quickly and efficiently. Topic updation is the process of identifying and renewing the discovered labels whenever the document clusters are updated dynamically. This paper focuses on topic updation model based on Testor theory. The proposed work is experimented using 20newsgroup and scientific literature data set. The experimental results demonstrate that the proposed algorithm exhibit better performance, compared to the existing algorithms for topic detection.
Correlated Concept based Topic Updation Model for Dynamic Corpora


Steinbach, M. , Karypis, G. , & Kumar, V. 2000. A Comparison of Document Clustering Techniques (pp. 1–2). International Conference on Data Mining: Knowledge Discovery and Data Mining (KDD) Workshop on Text Mining.


Index Terms

Computer Science

Information Sciences

Keywords

Document Clustering; Static Clustering; Dynamic Clustering; Topic detection; Topic updation; Testor Theory; F-Measure and Purity.