Abstract

Text mining has becoming an emerging research area now-a-days which helps in extracting the useful information from large amount of natural language text documents. The necessity of grouping the documents for different applications is gaining comprehensive review of the techniques used to improve the efficient time consumption, challenges, research issues are presented. The techniques presented in the review are k-means clustering, fuzzy c means clustering, support vector machine classifiers, naive Bayes classifier, Hidden Markov Model (HMM). Furthermore, discussion of the advantages and disadvantages of each technique is contributed to a better understanding and compared with the existing techniques based on the efficiency and computational time.

References


Index Terms

Computer Science  Information Systems

Keywords

Clustering, text mining, k-means clustering