A Topic-driven Summarization using K-mean Clustering and Tf-Isf Sentence Ranking

International Journal of Computer Applications
© 2013 by IJCA Journal

Volume 79 - Number 8
Year of Publication: 2013

Authors:
Rajesh Wadhvani
R. K. Pateriya
Devshri Roy

10.5120/13764-1608

Abstract

Enormous online information is available due to the World Wide Web. This needed efficient and accurate summarization systems to extract significant information. Text summarization system automatically generates a summary of a given document and helps people to make effective decisions in less time. In this paper two methods have been proposed for query-focused multi-document summarization that uses k-mean clustering, term-frequency and inverse-sentence-frequency method for sentence weighting to rank the sentences of the documents with respect to a given query. The proposed method finds the proximity of documents and query, and later uses this proximity to rank sentences of each document. It is assumed that the document which is nearer to a query might contain more meaning full sentences with respect to the information need expressed by user's query further if a sentence contains rare query term than it is more informative than the sentences that contains frequent query term. Both methods first gives weights to documents according to their proximity and use these document weights to rank each of their sentences with tf-idf ranking function. A relative study for proposed methods has been done and experimental results shows that both methods are comparable because of a slight difference in performance. DUC
2007 test dataset and ROUGH-1. 5. 5 summarization evaluation package is used for evaluation purpose.

References


- Laszlo Grunfeld, Kui-Lam Kwok, "Sentence Ranking Using Keywords And Meta-Keywords," Publisher Springer Netherlands, Volume 32, pp 229-258,2006.


gov/projects/duc.

Index Terms

Computer Science  Artificial Intelligence

Keywords

Sentence Extraction  Document Clustering  F-score