Abstract

Large number of techniques for keyword extraction have been proposed for better matching of documents with the user’s query but most of them deal with tf-idf to find the weight age of query terms in the entire document but this can result in improper result as if a term has a low term frequency in overall document but high frequency in a certain part of the document then that term can be ignored by traditional tf-idf method. Through this paper, the keyword extraction is improved using a hybrid technique in which the entire document is split into multiple domains using a master keyword and the frequency of all unique words is found in every domain. The words having high frequency are selected as candidate keywords and the final selection is made on the basis of a graph which is constructed between the keywords using Word Net. The experiments, conducted on various documents show that proposed approach outperforms other keyword extraction methodologies by enhancing document retrieval.

References
1. Information Retrieval Research, Jonathan Furner, School of Information and Media Studies, and David Harper, School of Computer and Mathematical Studies, The Robert Gordon University, Aberdeen, Scotland. (Eds)

2. Important problems in information retrieval, Dagobert Soergel College of Library and Information Services University of Maryland College Park, MD 20742

3. "Keyword extraction—a review of methods and approaches" Slobodan Beliga University of Rijeka, Department of Informatics Radmile Matejčić 2, 51 000 Rijeka, Croatia

4. Effective Approaches For Extraction Of Keywords Jasmeen Kaur, Vishal Gupta, ME Research Scholar Computer Science & Engineering, UIET, Panjab University Chandigarh, (UT)-160014

5. Understanding Inverse Document Frequency: On theoretical arguments for IDF, Stephen Robertson Microsoft Research 7 JJ Thomson Avenue Cambridge CB3 0FB UK

6. Keyword Extraction using graph based approaches, R. Nagarajan, Dr. S. Anu H Nair, Dr. P. Aruna, N. Puviarasan Department of Computer Science & Engineering, Annamalai University, Tamilnadu, India


A Statistical Approach of Keyword Extraction for Efficient Retrieval

Processing” in Proc. of 12th Int. Conf. on Asia-Pacific Digital Libraries, ICADL 2010, Gold Coast, Australia, LNAI v.6102, pp. 102-111, 2010


22. Slobodan beliga, University of Rijeka, Department of Informatics Radmile Matejčić 2, 51 000 Rijeka, Croatia,"Keyword extraction a review of method and approaches"


24. "Domain keyword extraction technique: A new weighting method based on frequency analysis" Rakhi Chakraborty ,Department of Computer Science & Engineering, Global Institute Of Management and Technology, Nadia, India


Index Terms

Computer Science Information Sciences

Keywords

Information Retrieval, Domain Splitting, Natural Language Processing, Inverse Document Frequency, Word Net