Abstract

In this paper, the use of TF-IDF stands for (term frequency-inverse document frequency) is discussed in examining the relevance of key-words to documents in corpus. The study is focused on how the algorithm can be applied on number of documents. First, the working principle and steps which should be followed for implementation of TF-IDF are elaborated. Secondly, in order to verify the findings from executing the algorithm, results are presented, then strengths and weaknesses of TD-IDF algorithm are compared. This paper also talked about how such weaknesses can be tackled. Finally, the work is summarized and the future research directions are discussed.

References


**Index Terms**

Computer Science

Data Mining

**Keywords**

TF-IDF, Data Mining, Relevance of Words to Documents