Abstract

The vector space model is one of the classical and widely applied information retrieval models to rank the web page based on similarity values. The retrieval operations consist of cosine similarity function to compute the similarity values between a given query and the set of documents retrieved and then rank the documents according to the relevance. In this paper, we are presenting different approaches of vector space model to compute similarity values of hits from search engine for given queries based on terms weight. In order to achieve the goal of an effective evaluation algorithm, our work intends to extensive analysis of the main aspects of Vector space model, its approaches and provides a comprehensive comparison for Term-Count
Model, Tf-Idf model and Vector space model based on normalization.

References


Index Terms

Computer Science

Data Mining

Keywords
Vector Space Model  Information Retrieval  Tf-idf  Term- Frequency  Cosine Similarity