Abstract

With the exponential growth of documents available to us on the web, the requirement for an effective technique to retrieve the most relevant document matching a given search query has become critical. The field of Information Retrieval deals with the problem of document similarity to retrieve desired information from a large amount of data. Various models and similarity measures have been proposed to determine the extent of similarity between two objects. The objective of this paper is to summarize the entire process, looking into some of the most well-known algorithms and approaches to match a query text against a set of indexed documents.

References


Index Terms
Keywords

Weighting Measures, TF-IDF, Cosine Similarity Measure, Jaccard Similarity Measure, Information Retrieval.