Abstract

Phrase query evaluation is an important task of every search engine. Optimizing the query evaluation time for phrase queries is the biggest threat for the current search engine. Usually, phrase queries are a hassle for standard indexing techniques. This is generally because, merging the posting lists and checking the word ordering takes a lot of time. This paper proposes a new technique called Triple Indexing to index web documents which optimizes query evaluation time for phrase queries by reducing the time for merging the posting lists and checking the word ordering. In addition, a proper procedure has been put forward for document ranking using an extended vector space model. The 4 Universities dataset and Industry Sector dataset of Carnegie Mellon University has been used for experimental purpose and it has been found that using the proposed method with a modern machine, the query time for phrase queries is reduced by almost 50 percent, compared to a standard inverted index.

References


Index Terms

Computer Science
Information Sciences
Keywords
Triple Index  Inverted Index  Query Optimization  Phrase Queries  Vector Space Model