Abstract

Query Expansion using Pseudo Relevance Feedback is a useful and a popular technique for reformulating the query. In our proposed query expansion method, we assume that relevant information can be found within a document near the central idea. The document is normally divided into sections, paragraphs and lines. The proposed method tries to extract keywords that are closer to the central theme of the document. The expansion terms are obtained by equi-frequency partition of the documents obtained from pseudo relevance feedback and by using tf-idf scores. The idf factor is calculated for number of partitions in documents. The group of words for query expansion is selected using the following approaches: the highest score, average score and a group of words that has maximum number of keywords. As each query behaved differently for different methods, the effect of these methods in selecting the words for query expansion is investigated. From this initial study, we extend the experiment to develop a rule-based statistical model that automatically selects the best group of words incorporating the tf-idf scoring and the 3 approaches explained here, in the future. The experiments were performed on FIRE 2011 Adhoc Hindi and English test collections on 50 queries each, using Terrier as retrieval engine.
References


Index Terms

Computer Science

Information Sciences
Keywords
Pseudo Relevance Feedback  Query Expansion  tf-idf  equi-frequency partition