Abstract

Various information retrieval models generate different ranking list as output. This paper presents the comparative analysis of the vector space model and the probabilistic model. Effect of stopword removal is also discussed. A new hybrid model is introduced that combines the Vector Space Model and the Probabilistic model. The resultant model gives better performance. For experiments, we have constructed English-Hindi IR test collection from EMILLE parallel corpus. Relational (stop) words are considered for improving the search results. F-measure and AIP (Average Interpolated Precision) are used for evaluation.

Reference

- G. Salton, A. Wong and S.S. Yang 1975. A vector space model for automatic indexing,
communications of the ACM, 18, pages 613-620.

Index Terms

Computer Science Information Retrieval

Key words

IR models comparison Stopword removal

English-Hindi parallel corpus

Relational Stopwords

Hybrid Model

Vector Space

Model
Probabilistic Model