Application of Topic Based Vector Space Model with WordNet

ABSTRACT
Topic Based Vector Space Model (TVSM) proposed a new vector space that its dimensions is composed of topics. Every term and document is represented by vectors inside this vector space. By using topics as dimensions TVSM tries to overcome word mismatch between terms with similar topics in finding relevant documents to query. This study proposes to develop relations between terms using WordNet and thesaurus to help TVSM calculating similarity between documents. Relations between terms are represented by relation score. This study proposes a way to find optimal relation score for a set of documents. To help indexing documents with multi language terms this study also proposes to use dictionary to expand query terms.

INDEX TERMS
Index Terms are available to subscribers and IEEE members.