Abstract

Information retrieval systems that use literal term matching between user query and keywords of a document can produce inefficient outputs, because the term matching does not extract knowledge or conceptual information from the document; therefore, the results of documents are retrieved based upon the matched keywords, but may be irrelevant to the concepts of documents. Another approach is the matrix oriented technique in which Latent Semantic Indexing (LSI) is applied. LSI can extract conceptual information underlying in documents. This technique changes the keyword representation of documents into a term-by-document matrix and uses matrix decomposition techniques to refine the matrix for the better representation of term-document relationship. The main purpose of this thesis is to compare between two matrix decompositions: singular value decomposition (SVD) and semidiscrete matrix decomposition (SDD), of matrix of Thai text-based information in four domains. The results show that the SDD occupied less storage but took more time for decomposition than did the SVD significantly. Precision and recall of both matrix decomposition methods depend on various setup values during processes of design and implementation and the objectives of information retrieval systems.