

Abstract

The growth of number and variety information for example documents, user more difficult looking for wanted information. When browsing documents using search engine, there are a lot of documents as results but sometimes the documents are not related to user's needs. User still must searching the documents one by one in the sense of meeting their needs. Hence text grouping the documents is needed by applying one of technique which related to Text Mining and Information Retrieval that is "Cluster Based Retrieval (CBR)". The implementation of CBR in search engine can make user searching the documents easier.

The CBR that is used is Cover Coefficient Clustering Method (C3M), a method applying partitional clustering having the character of assign documents exactly in one cluster (non-overlapping) by calculating cover coefficient to determine the number of clusters and seedpower. C3M also use probabilistic models and terms similiarity to show the relationship of documents and term queries.

Examination performed to analyze the results of grouping C3M Algorithm based on parameter cluster quality using silhouette coefficient and analyze the causes of clusters quality. Cluster quality of C3M Algorithm depends on the relevance documents that retrieve, the number of cluster, the number of term query and the number of hitlist documents. The implementation of CBR using C3M Algorithm can show a good cluster quality by the result of silhouette coefficient that always in positive value.

Keywords: cluster based retrieval, cover coefficient clustering algorithm partitional clustering, silhouette coefficient