

Abstract

Nowadays, when searching documents, the search result will sort retrieved documents based on their rank. The results sometimes irrelevant and different from user's expectation. One alternatives to improve the search results is to clusterize it.

Documents in this final project will be in Indonesian language with static amounts and using *Descriptive K-Means* algorithm. The difference between original K-Means and Descriptive K-Means(DKM) is that DKM have a certain phases to get labels and to allocate documents.

After implementation, this algorithm produce clusters that contains relevant documents to cluster label due to its query that uses AND relation to allocates documents, it means that every term label must be in every documents in the cluster.

This algorithm become uneffective if user queries are too few in the document collections. If condition like this happen then some of clusters will contains less then minimal required amount of documents.

Based on *precision* and *recall* value, the most effective search result comes when K is set =2. This values decrease when K value gets bigger. *Convergence* value cannot be used to conclude a cluster is good or not because its really depends on user's queries and amount of documents within the cluster.

Keywords: *Convergence, Descriptive K-Means, precision, recall.*