

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 alternative to improve the search results is to clusterize it.

Documents in this final project will be document collection in Indonesian language and be clustered using LINGO algorithm. LINGO is clustering algorithm which ensure that both contents and description (labels) of the resulting groups are meaningful to the users.

After implementation, this algorithm produce clusters that contains relevant documents to cluster label due to for each document is allocated to the cluster based on it's similarity to the label cluster.

To determine the labels to describing cluster, this algorithm will check the occurrence of the term and complete phrase in the documents. So the algorithm will become effective if processed documents contains recurrent topic terms, otherways it will become ineffective if topic terms or phrases in the documents are interpreted in many different terms.

Due cluster label testing process using precision and recall on Term Frequency (TF) weighting and Term Frequency – Inverse Document Frequency (TF-IDF) generate good result for the both weighting methods.

Keywords: *cluster, clustering, LINGO, complete phrase, precision, recall.*