

Abstract

The growth of variety and number of document that give impact in file searching process. File searching which available now just appear the sequence searching result based on level of compatibility (document ranking). The result were appeared occasionally inappropriate (inrelevant) with that was wanted by the user.

One of the methods of facilitating the user in looking for the document was by using categorizing of the document. With the existence of the name of the category from each cluster could facilitate the user because the name of the category could reflect the contents from a document collection. One of the methods of grouping the document was with clustering. In this final project will be carried out cluster the Indonesian document with the K Means algorithm. Before clustered was needed pre processing process that was case folding, parsing, the disposal stopword and stemming. This process was needed to reduce the number of words that was processed. The name of the category was given to each cluster. The naming of the category was counted from the frequency the of the words from each cluster.

The analysis that have done in this final project is compare two methods weight and the distance correlation between the object. Moreover, the analysis done to know the compatibility of the name of the category with document in each cluster and analysed results cluster that was produced by seeing the value convergence, precision and recalled.

Keywords : name of category, k means, *clustering*, document, *pre processing*