

Abstract

The rapid usage of internet as an information spreading medium causes information availability in the large amount, various, and commonly unstructured. This lead to the need of information management to get useful knowledge. One of important components in information management is grouping text by categorization and clustering so this can help people to get information according to certain category fastly and accurately.

In this final project, used concept based method (statistical concept based mining model and modified statistical concept based mining model) and traditional method like TF IDF for weighting terms. The output from this process will be input for categorization and clustering Indonesian news articles. After that, will be measured the categorization and clustering performance.

Concept based method have high performance and varied in some categories. From this research result also can be known that concept based mining model has better accuration than TF IDF method. It shows us that concept based mining model can be a weighting term method in order to get good categorization and clustering performance.

Key Word: statistical concept based mining model, modified statistical concept based mining model, TF IDF, categorization, clustering