Abstract

As a rich with vocabulary language, Indonesian language has many words with the same meaning (synonym). This can cause news report being grouped in a non relevant category with the news' content. Therefore, a method to to process data is needed for getting te benefit from that data. One of the method used to process news is data mining. In data mining, there is a method that is used often, which is clustering. Clustering is the grouping of object according to its characteristic. The news grouping can use the clustering methode with the purpose to group a news article appropriate with its news topic.

In this final assignment, a clustering method is implemented, which is the Clustering based on Frequent Word Sequences (CFWS) algorithm on Indonesian language news article. CFWS is an algorithm that represents documents by using the most frequent word sequncesthat appear inthe document. By using this algorithm, the dimension of the document can be reduced significantly so the clustering process can be more efficient. THe testing was done to see the quality of the final cluster according to the accuracy calculation with F-Measure.

According to the test that have been done, CFWS algorithm produce a good quality of cluster. Beside that, the CFWS algorithm can produce a good cluster for the data set with a similar topic and different topic.

Keyword : *data mining, clustering, CFWS, F-measure*

.