

Abstract

With the rapid increase of information / documents circulating on the internet making it possible for a document can be grouped into two or more categories. For that it needed a method to be able to classify documents in which a document can be grouped into two or more categories.

Overlapping Clustering Cover Coefficient Method (OC3M) is a method of grouping documents with probabilistic models, terms similarity, and the seed documents as early initialization of cluster formation. In this method, it is applied an overlapping properties, which is causing a document to be able occupy more than one cluster.

A tests conducted in this thesis in clustering a document with OC3M algorithm is by analyzing a produced clusters using Silhouette Coefficient value and analyzing the things that affect the quality of the formed cluster. The quality of clusters is influenced by the number of documents used, type of document, the document similarity with the cluster's center, and also influenced by the overlapping coefficient which is a parameter that determines a similar documents could be grouped into a different clusters. The result of the experiment in clustering using OC3M algorithm is quite good, it's shown in the silhouette coefficient value, which is a positive value.

Keywords : *OC3M, Overlap, Overlapping, Clustering, Cluster*