

Abstract

Along with the growing technology, there will be a rapid growth in Document diversities that causes degrading impacts on the performance of current search engines. A lot of ir-Relevant Documents will be retrieved and served to the user. Therefore, it will be troublesome for the user to find the documents that he wants.

Clustering is one of the solutions that can be used to solve this problem. Clustering, itself, is a process of collecting similar Documents into the same clusters. In this final project, clustering is done by using Buckshot algorithm. Buckshot is a method that combines HAC (Hierarchical Agglomerative Clustering) and K-Means algorithm. This method makes use of HAC algorithm to determine the initial centroids in order to make better quality clusters compared to the one that uses K-Means algorithm only.

The analysis done in this final project is meant to compare the quality of clusters formed using Buckshot and K-Means algorithm. The quality of the clusters will be determined by using Cohesion, Separation, and Cohesion/Separation values. During the test the Buckshot Algorithm show better result compare K-Means when the user pick the number of clusters that's its defferent from random sample used by the system. The Quality of the clusters is also influenced by precision and recall values.

Keywords : *Clustering, Search engine, Buckshot Algorithm, HAC Algorithm, K-Means Algorithm, Cohesion, Separation, Precision, Recall.*