

Abstract

News today is not only disseminated through electronic media and print media, but also disseminated through the internet. So the amount of news available is enormous and it took a long time to group them manually. Clustering can be used as a solution to these problems. One method that can be used for word document clustering is a method Hierarchical Agglomerative Clustering (HAC). In this final project HAC method used is Single link, Complete link, and Average link.

HAC method has the ability to combine data to create hierarchies, which have similar data will be placed in an adjacent hierarchy and that it bears no resemblance to the hierarchy that placed far apart. Each document will be considered as a cluster, then combined according to the HAC method used, repeated until the number of clusters as desired.

Hierarchy that is formed will be calculated using the cophenetic correlation coefficient of quality, while the quality of the formed clusters will be calculated using silhouette coefficient and purity.

Keywords : clustering, cophenetic correlation coefficient, silhouette coefficient, purity