

Abstract

Documents of the same category in large numbers is difficult to distinguish the similarity between documents with the other documents. One way that can be used to overcome this problem is to document clustering. For the number of clusters, users do not even know how much the right to perform clustering documents. So, It required a clustering method which can produce the number of clusters automatically. One of the many methods that can be used is Hill Climbing.

Hill Climbing will identify the movement of variants of each stage of the formation of clusters and analyze the pattern to find the optimum global so that the number of clusters will be formed automatically. For the clustering method used is one of the Hierarchical Agglomerative methods, namely Centroid Linkage Hierarchical Method (CLHM). Each document will be considered as a cluster, then combined with algorithms CLHM repeated until the number of clusters as desired. The results of the cluster that is formed will be calculated that quality with the evaluation parameters and purity values.

Keywords: Clustering, CLHM, Hill Climbing, variance, purity