Abstract

The growing up of information technology, couse of the number of digital text documents also grow up rapidly. To retrieve of information about current event related to specific topics in a documents collection will be more difficult. *Cluster*ing is one of data mining method which is unsupervised learning to classify documents based on *similarity*. One of *cluster*ing method is single pass *cluster*ing. Single pass *cluster*ing is a type of *cluster*ing algorithm that try to create group of data one by one and the formation of the group performed in line with the evaluation of any data entered into the *cluster* process. The number of *cluster*s generated is depended on the threshold value. The *similarity* between the documents uses the standard cosine *similarity*.

In this final project, the quality of the *cluster*s result is measured by two parameters cohesion and separation. Value of 0.03 and 0.029, can produce the average of cohesion 0.13863266 and the average of separation 0.009379814.

Keywords: ***single pass clustering, unsupervised learning, cosine similarity, threshold, cohesion, separation.***