

Abstract

The growth of number and variety information for example news article, consumer more difficult looking for wanted information. Hence text grouping the article is needed by applying one of technique which related to Data Mining, that is “ clustering ”.

The clustering method that is used is Ward’s method, a method applying principle of agglomerative hierarchical clustering having the character of bottom up where grouping object is formed hierarchyly started from bottom by determining every object form each cluster till to up that is grouping the most similar couple level by level until all element object in one cluster. The most similar couple is couple whose fusion results in the most minimum increase of error sum of square.

Examination performed to analyze result of grouping Ward’s method based on parameters of error sum of square(ESS), Cophenetic Correlation Coefficient(CPCC), FMeasure, and execution time. Addition of number testing document and topic in the collection cause changing ESS tend more increase. While addition of number testing document also influences value of CPCC, the highest value is got by the collection that has the fewest number (consist of 10 document) that is 0,9519(quality of cluster hierarchy is 95,19%). While based on FMeasure and execution time, average of the highest FMeasure indicates accuration value in grouping obtained by complete linkage which is as comparison method. eventhough Ward’s method is quicker in the grouping seen from the execution time.

Keywords : grouping document, agglomerative hierarchical clustering, Ward’s method, error sum of square, cophenetic correlation coefficient, FMeasure.