

ABSTRACT

A collection of many documents would be difficult to know the characteristics of the information. Need for a system that can divide that many document into a little collections based on the characteristic. Clustering is the grouping of an object based on the information residing on those object. The way is by classify the same information characteristic into the same class and an object with the different information characteristic into the other class. Therefore in this research will be made clasterisation using Fuzzy C-Means (FCM). FCM is a clustering technique in which the existence of each data point in a cluster is determined by the degree of membership. In the grouping of documents by using this method is need process lowercasing, removing punctuation and numbers, tokenization, stopwords removal, stemming and term weighting. On the stemming, the algorithm used is Arifin-Setiono stemming algorithm. For term weighting the method used is tf-idf term weighting. FCM using fuzzy clustering models that can make data into member from all class with the different degree of membership between 0 and 1.

Keywords : *Clustering, documents, Arifin-Setiono stemming algorithm, tf-idf term weighting, Fuzzy C-Means.*