# INDONESIAN DOCUMENT CLUSTERING USING

# FUZZY  C-MEANS

**Isna Mariam, Julio Adisantoso**

Department of Computer Science

Faculty of Mathematics and Natural Sciences

Bogor Agricultural University

julioipb@gmail.com

2011

## Abstract

Document clustering enables a user to have a good overall view of the information contained in the document. Most classical clustering algorithms assign each data to exactly one cluster, thus forming a crisp partition of the given data. Recently, fuzzy clustering approach allows for degrees of membership, to which a data belongs to different clusters. Document clustering aims to make a cluster that is internally coherent but clearly different from other clusters. The documents that are used in this research are several documents from journal of horticulture and documents of medical plantations. All documents in the collections are clustered by using fuzzy C-Means algorithm. Furthermore, in this research threshold is used to weight the words that engage in the clustering process. The appropriate uses of threshold may give a better accuracy for the clustering result. The best result in this research is obtained when the threshold value is 1.5 and fuzzifier value is 2 for the documents from journal of horticulture, whereas for the documents of medical plantations the best result is obtained when the threshold value is 0.75 and fuzzy coefficient value is 2.