Abstract

Research in the field of text mining to process entries or words from the Qur'an is very beneficial for Muslims. This research aims to establish grouping synonym sets for thesaurus in the words of the Qur'an. This research is used because the source of knowledge about the science of the Qur'an is still lacking. The dataset in this study uses Corpus Qur'an and English Translation. This research is a development study of an article that has been published, namely "Development of the Synonym Vocabulary of the Qur'an Set with the WordNet Approach" by Laras Gupitasari. Word input from this research system uses nouns from English words in the Koran. The output of the system produces several groups that have the same level of closeness of meaning, the first group means the word in the group has a close meaning. To produce output, this study uses word grouping with a hierarchical clustering algorithm and calculates the distance using Path Similarity, then groups the results according to the meaning of the word entry. The evaluation in this study produced a 76% F-Measure , the F-Measure Value is an evaluation to measure the accuracy of predictions issued by the system.

Keywords: Thesaurus, Synonym Sets, Al-Qur'an, Text Mining, Hierarchical Clustering