ABSTRACT

PASKIANTI, KRISTINA. Document Classification using KNN Fuzzy Algorithm. Supervised by YENI HERDIYENI.

Many documents about the medicinal plants are not fully utilized because of difficulty finding information. Computational techniques in the form of search engine may be one solution to solve the issue. This research proposed a new system for document classification and information retrieval system of medical plants. Chi-Square is used to extract term that represents a specific category. This terms are called features. KNN Fuzzy algorithm is used to classify documents of medical plants. Document classification done for two categories, there are family and disease. The experiment result show that number of features produced in significant level of 0.01 is greater than number of features produced in significant level of 0.001. This makes the level of accuracy produced in significant level of 0.01 is higher than significant level of 0.001. KNN Fuzzy algorithm is effective for document that has uniform characteristic and document that has uneven distribution. According to experiment result, level of accuracy achieved in family model is 94.87% and level of accuracy achieved in disease model is 100%. Average precision of information retrieval for single query is 0.96 and for double query is 0.89.

Keywords: document classification, KNN Fuzzy algorithm, information retrieval, Chi-Square feature selection