ABSTRACT

YOGA HERAWAN. Feature Extraction of Medicinal Plants using Chi-Square with Naïve Bayes Classifier. Supervised by YENI HERDIYENI.

This research presented a system for extracting terms and classifying medicinal plants documents using chi-square and naïve bayes classifier. Term extraction technique was used to make the classifier work efficiently and to increase classification accuracy. The criteria used in this research were the family of medicinal plants and utilization of medicinal plants for medication.

The classification results were used to build an information retrieval system of Indonesian medicinal plants. This research used two significance levels for generating critical value, i.e 0.001 and 0.01. The experiment result showed that the critical value using significance level of 0.001 has better accuracy than the critical value using significance level 0.01. Accuracy of classification system using significance level of 0.001 were 97.44% for family and 89.74% for utilization of medicinal plants criteria. The information retrieval system tested using 29 queries about family and utilization of medicinal plants. The information retrieval system had an average value generated was 93.26%.

Keywords: document classification, naïve bayes classifier, chi-square, feature selection, information retrieval.