

Abstract

Increasingly rapid technological developments led to the amount of data and documents, such as news documents were greater in number. Large amounts of search data causes difficulty to find data that fits document seeker needs. One of Data Mining task, Text Categorization, is one solution to this problem.

Text Categorization is the task of automatically classifying unlabelled natural language documents into certain categories. In text categorization, term weighting method is such an important step to improve the effectiveness of text categorization, for different terms have different importance in a text. The system was built aimed to compare two methods of term weighting, TF • IDF and TF • Chi-Square. After getting the weighting, then the system performs the classification of the document is entered.

System performance is measured by the method of SVM through the accuracy of the classification results are affected by the weighting of the two methods. Evaluation parameters used are accuracy, recall, precision, and F-measure. The experimental results prove that the weighting term using TF • Chi-Square gives a better performance than the TF • IDF. Performance is also influenced by the size of threshold and the number of documents used.

Keywords : *Text Categorization, Term Weighting, TF • IDF, TF • Chi-Square, recall, precision, F-measure*