Abstract

Technological developments lead to the accumulation of data in the form of text documents either online or offline. Text documents that accumulate causing difficulty to find documents as needed. For ease of searching the appropriate documents, it is required to accompany the keywords document. Keywords representing the contents of the document as a whole, so that readers can easily search for documents in accordance with the required.

In this final project, implemented Naive Bayes method to extract the document keywords. This process requires training documents that have been accompanied by keywords, so by using certain features can provide the system of learning how the probability or the chance to document the inputs into the keyword.

Tests performed to determine the accuracy of the keywords generated by the system using Naive Bayes method based on the parameters of precision, recall, and f-measure. The addition of training documents, increase in the accuracy of the results of keyword extraction. However, the use of the number of training documents is too large causes a decrease in the value of accuracy. Use of the 4 feature extraction processes yield better accuracy than the use of feature 2. Naive Bayes capability in extracting the correct keywords can be seen from the value of recall. In the span of 20 the number of keywords that generated the system, Naive Bayes can provide a recall value of 0.75 (about 75% of the system is capable of extracting keywords according to the input document). Extraction of keywords by adding stopwords elimination yield better accuracy than without stopwords elimination.

Keyword: Keyword extraction, keyword, naive bayes