Abstract

Along with the development of the internet, the more web pages are available in cyberspace. Many people are looking for the information about the health of the Internet because health is essential in human life. Users will find it difficult to get web pages quickly, solid and in accordance with the relevant health topics. Focused crawler can provide web pages in accordance with the topic in question. Focused crawler is to selectively browse and retrieve relevant web pages. Of course, in addition to crawler component, there is also to build a classifier focused crawler.

In this final task is the implementation and analysis of the focused crawler using Naive Bayes Classifier (NBC). The order of the browsing queue use the breadth-first search (BFS.) it is expected to use a Naive Bayes classifier accuracy of the results to obtain higher classification. Before testing the focused crawler, testing a number of web pages involved in the process of training and use of stopword or no training data. The goal is to produce a classification accuracy of the maximum results.

Shown in this final task that focused crawler using Naive Bayes classifier to produce the classification results of the web pages is relatively less because of the decrease in accuracy at every level.

Keywords: focused crawler, naive bayes classifier, classification, web pages, health, training.