Abstract

The rapid growth of world-wide web followed by the increase on demand of information becomes a challenge which never occurs in the past for general-purpose crawler and search engines. Search engines, namely google, yahoo! Alavista and etc., have been introduced and utilized to ease the user browsing information on the internet. Web crawler (crawler) is an automatic program that processes a web page into a search engine on which commonly used nowadays. Nevertheless, due to the abundant presence of web page, search engine and crawler frequently cannot lead to a maximum result. As a result, Focused Crawler is developed. Focused crawler will automatically download a web page that are matched with desired topic and carefully decides which URL to be scanned on the basis either in order or previous download history. In this final year project, the subject matter to be highlighted is music web.

Focused crawler requires classifier to distinguish between relevant and irrelevant webs, in which in this final year project, the classier used is Naïve Bayes. All relevant web pages will be extracted in terms of outgoing-link and it will be saved into frontier. Link will be crawled by employing algorithm method/ formula. Algorithm method that is suitable will affect web crawler efficiency. And the author decided to use Learning Anchor Algorithm.

Based on the implementation, the resulting of best accuracy is 100% when the testing link is http://gigsplay.com/ *with dataset 100 music and 100 nonmusic. While the lowest accuracy is 86.7% when the dataset is 300 music and 200 nonmusic at testing link* http://musik.kapanlagi.com/**.**

Keywords: focused crawler, sport web, naïve bayes, Learning Anchor Algorithm.