

Comparison of Three Crawling Libraries for Providing Herbal Information Resources

Vincentius Riandaru Prasetyo
Department of Informatics Engineering
University of Surabaya
Surabaya, Indonesia
vincent@staff.ubaya.ac.id

Risqi Wahyu Permana
Department of Informatics Engineering
University of Surabaya
Surabaya, Indonesia
s160419081@student.ubaya.ac.id

Oeke Yunita
Faculty of Pharmacy
University of Surabaya
Surabaya, Indonesia
oeke@staff.ubaya.ac.id

ABSTRACT

The growth of the herbal market and its potential worldwide is closely related to the need for information about herbal plants. Various companies use this information to develop herbal products, one of which is PT Saka Farma Laboratories (Kalbe Consumer Health). This information is obtained through public websites such as knapsackfamily.com, phytochem.nal.usda.gov, and ntbg.org. However, gathering this information takes a long time. One way that can be used in order to obtain information quickly is to use a crawling technique. Some of the crawling libraries often used are beautiful soup, selenium, and simple HTML dom. This study aimed to evaluate the crawling results of the beautiful soup, selenium, and simple HTML dom libraries. The websites used as crawl targets are knapsackfamily.com, phytochem.nal.usda.gov, and ntbg.org. Evaluation is done by comparing the crawling results of the libraries used. Based on the results of the tests, the Beautiful Soup library is ideal for crawling the three targeted herbal information websites. However, the selenium library and simple HTML dom can get all the information from the website phytochem.nal.usda.gov, while beautiful soup only gets some information.

Keywords—herbal information, crawling, beautiful soup, selenium, simple_html_dom