ABSTRACT

Methods Web Scraping has many diverse methods. In doing Web Scraping, the main focus is to identify the website, whether it is a statistical or dynamic website and also the data structure of the website. With the many methods of Web Scraping, this makes it quite difficult for users to choose the method that suits their needs, especially in this day and age where the need for information is needed quickly and precisely.

This final project aims to compare several methods of scraping data from several different characteristics of the web and to compare several parameters to find the appropriate and appropriate method in handling each of the characteristics of the site.

The results of this study take data on Simple Web as a sample for website statistics and Kompas and Tokopedia as a sample for website dynamics with 20 times testing for each parameter. In this study using 3 methods, namely Regular Expression, Parsing DOM and XPath. In this study using the Regular Expression method and get an average CPU usage of 3.57%, Memory of 5.69 KB and time it takes 1.94s. With the Parsing DOM method the average CPU usage is 2.63%, Memory is 388.04 KB and the time it takes is 2.76s. Then using the XPATH method the average CPU usage is 2.9%, Memory is 1,41 KB and the time it takes is 2.32s..

Keywords: *Web Scraping, Parsing DOM, XPath, Regular Expression, Dynamic Web Content*