## ABSTRAK

Pengumpulan data sudah menjadi berbagai kebutuhan pada saat ini, terlebih banyaknya sumber data di internet yang beragam. *Data extraction* atau proses pengambilan data dari internet dikenal juga dengan sebutan web scraping. Beberapa teknik *web scraping* diantaranya: *CSS Selector, HTML DOM, Regular Expression (Regex)* dan *XPath*. Banyaknya jumlah data yang bersebaran di internet akan cukup memakan waktu bila dilakukan web scraping dalam skala besar. Hadirnya berbagai inovasi baru menjadi sebuah solusi yang dapat digunakan, dengan adanya sistem paralel sebuah pekerjaan dapat diselesaikan dengan lebih cepat dan sistem paralel tersebut dapat dilakukan oleh *multiprocessing*. Penelitian ini bertujuan untuk mengetahui kinerja metode *web scraping* dengan implementasi *multiprocessing* didalamnya. Pengujian dilakukan terhadap masing-masing metode dengan cara melakukan sebuah proses *scraping,* kemudian diukur kinerja dari proses tersebut dan dibandingkan. Jumlah objek data yang didapat, penggunaan *CPU*, penggunaan memori, waktu proses dan penggunaan *bandwith network* dijadikan parameter pengukuran dalam percobaan. Hasil percobaan menunjukkan*XPath* 10,31% lebih cepat dalam melakukan proses dan memiliki perbandingan hasil yang tidak terlalu jauh pada parameter penggunaan memori dengan score *351.642 bytes* untuk *XPath* dan *334.438* untuk *CSS Selector*.



Kata Kunci: *CSS Selector, HTML DOM,, Multiprocessing, Regex, Web Scraping , XPath.*

## *ABSTRACT*

*Data collection has become a necessity at this time, especially the large number of data sources on the internet that can be used for various needs. Data extraction or the process of retrieving data from the internet is also known as web scraping. Several web scraping techniques include: CSS Selector, HTML DOM, Regular Expression (Regex) and XPath. The large amount of data that is scattered on the internet will be quite time consuming if web scraping is carried out on a large scale. The presence of various new innovations is a solution that can be used, with the parallel system a job can be completed more quickly and the parallel system can be done by multiprocessing. This study aims to determine the performance of the web scraping method with the implementation of multiprocessing in it. Tests are carried out on each method by carrying out a scraping process, then the performance of the process is measured and compared. The number of data objects obtained, CPU usage, memory usage, processing time and network bandwidth usage are used as measurement parameters in the experiment. The experimental results show XPath is 10,31% faster in processing and has a comparison of results that is not too far on the memory usage parameter with a score of 351,642 bytes for XPath and 334,438 for the CSS Selector.*



*Keywords: CSS Selector, HTML DOM, Multiprocessing, Regex, Web Scraping, XPath.*