ABSTRACT

Web Scraping is the process of retrieving a semi-structured document from the internet, generally in the form of Web pages in a markup language such as HTML (Hypertext Markup Language) or XHTML (Extensible HyperText Markup Language), the data is analyzed to retrieve certain data from the web for use. for other purposes. The number of web scraping actors due to the higher information needs and the desire to obtain information quickly makes many people who want to use and learn web scraping techniques to be able to collect information from an information provider website.

This study aims to collect all methods that can be used to protect data on websites from data scraping techniques and bots and also conduct direct tests from bots to prove the method. The research will be conducted on a website made by the author's team which will implement several methods for website protection such as javascript checker, inspect element block, CAPTCHA, javascript feature management, ip session management, data traffic management, and blocking suspicious IPs.

The parameter that will be used as a benchmark for the level of security that is good or not is how this method can prevent a bot from being able to enter the information provider website system that will be created by the author which will then be tested using bot scraping.

Keywords : *Web Scraping, How to stop web scraping, Data Collectors, Search Engines, Search Engines, CAPTCHA, Web crawlers, Online Scraper Sites.*