

ABSTRAK

Web Scraping adalah proses pengambilan sebuah dokumen-semi terstruktur dari internet, umumnya berupa halaman-halaman Web dalam bahasa *markup* seperti HTML (*Hypertext Markup Language*) atau XHTML (*Extensible HyperText Markup Language*), data tersebut dianalisis untuk diambil data tertentu dari web tersebut untuk digunakan bagi kepentingan lain. Banyaknya pelaku *web scraping* dikarenakan kebutuhan informasi yang semakin tinggi dan keinginan untuk memperoleh informasi dengan cepat membuat banyak orang yang ingin menggunakan dan mempelajari teknik *web scraping* untuk dapat meraup informasi dari sebuah situs web penyedia informasi.

Penelitian ini bertujuan untuk mengumpulkan seluruh metode yang dapat digunakan untuk memproteksi data pada situs web dari teknik scraping data dan *bot* dan juga melakukan test secara langsung dari bot untuk pembuktian metode. Penelitian akan dilakukan pada sebuah situs web buatan tim penulis yang akan diimplementasikan beberapa metode untuk proteksi situs web seperti javascript checker, blokir inspect element, CAPTCHA, manajemen fitur *javascript*, manajemen *ip session*, manajemen trafik data, dan blokir *IP* yang mencurigakan.

Parameter yang akan dijadikan sebagai tolak ukur untuk tingkat keamanan yang baik atau tidaknya adalah bagaimana metode tersebut dapat menghalangi sebuah *bot* untuk dapat masuk kedalam sistem situs web penyedia informasi yang akan dibuat oleh penulis yang kemudian akan dilakukan testing menggunakan *bot scraping*.

Kata Kunci : *Web Scraping, How to stop web scraping, Pengumpul Data, Mesin Pencari, Search Engine, CAPTCHA, Web crawler, Pencakar Situs online.*