

ABSTRAK

Semua jenis informasi sangat mudah didapatkan dalam dunia modern saat sekarang ini. Salah satu cara untuk memperoleh informasi yang ingin didapatkan yaitu dengan mencari atau *searching* di dunia maya lewat akses internet. Jumlah informasi yang ada pada saat *searching* sangat banyak dan tidak terstruktur (*big data*). Untuk mengumpulkan dan memilah informasi dari *big data* tersebut secara manual membutuhkan waktu dan usaha yang tidak sedikit. Sehingga dibutuhkan cara yang lebih efektif untuk mengumpulkan data yaitu dengan *Web Scraping*.

Web Scraping adalah teknik mengambil dan mengekstraksi data pada sebuah web tertentu dan mengumpulkannya pada satu file. Dengan cara ini, data yang ingin dicari akan lebih mudah untuk didapatkan. Data yang diperoleh juga sudah terstruktur sehingga dapat memudahkan pencari data dalam membaca informasi yang diperoleh.

Dalam penelitian ini, pengerjaan dilakukan menggunakan teknik *Web Scraping* dengan metode *HTML parsing* pada *Google Scholar* untuk menemukan data artikel ilmiah yang akan dicari. Data yang telah dicari akan dikelompokkan dan disusun kedalam sebuah file dengan format *Comma Separated Values (CSV)*. Dengan *output* berupa file *csv* ini diharapkan dapat mempermudah pencari informasi untuk mendapatkan, mengumpulkan, mengolah dan menganalisis data artikel ilmiah yang mereka cari.

Kata Kunci: *Web Scraping, Google Scholar, Big Data, Automatic Data Collection, Scientific Article.*