

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dunia maya memegang peranan penting dalam era teknologi informasi dan digital saat sekarang ini. Jumlah informasi dan data banyak bertebaran di internet dari berbagai jenis sumber seperti web, *blog*, *wordpress* dan lain sebagainya sehingga terciptalah istilah *big data* [1]. Untuk mendapatkan dan mengumpulkan data tertentu yang diinginkan dari *big data* tersebut secara manual (*copy paste*) tidak mudah, dibutuhkan usaha dan waktu yang tidak sedikit.

Salah satu cara untuk memudahkan pengumpulan informasi dan data yaitu dengan menggunakan teknik *Web Scraping* [2]. Cara ini dapat memproses data yang diinginkan pada suatu web dan mengumpulkannya pada satu file [3]. Data yang akan didapatkan bisa dibentuk dalam sebuah struktur sehingga lebih memudahkan dalam membaca informasi yang di peroleh [4].

Pada penelitian ini, penulis akan mengumpulkan data terkait informasi artikel ilmiah pada pencarian Google Cendikia (*Google Scholar*). Penulis akan mengaplikasikan teknik *Web Scraping* dengan metode *HTML parsing* menggunakan bahasa pemrograman Python. Informasi yang diperoleh akan dirangkap otomatis ke dalam suatu file dan dibentuk secara terstruktur sehingga memudahkan dalam pembacaan data.

1.2 Rumusan Masalah

Berdasarkan pokok permasalahan diatas, penulis menggali rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana mengumpulkan data artikel ilmiah di *Google Scholar* menggunakan teknik *Web Scraping* dengan metode *HTML parsing*?
2. Seberapa efektif penerapan teknik *Web Scraping* dengan metode *HTML parsing* dalam pengumpulan data artikel ilmiah di *Google Scholar*?

1.3 Tujuan dan Manfaat

Penelitian ini bertujuan untuk mendapatkan dan mengumpulkan data informasi artikel ilmiah di *Google Scholar* ke dalam suatu file secara otomatis dan terstruktur menggunakan teknik *Web Scraping* dengan metode *HTML parsing*. Keakuratan dalam mendapatkan dan mengumpulkan data serta kestrukturannya menjadi tolak ukur kesuksesan pada penelitian ini. Implementasi menggunakan bahasa pemrograman Python dengan harapan akan menghasilkan informasi yang jelas dan efektif serta waktu yang singkat.

1.4 Batasan Masalah

Untuk lebih memfokuskan penelitian Tugas Akhir ini, maka diberikan beberapa batasan masalah sebagai berikut:

1. Metode *Web Scraping* yang digunakan dikhususkan pada *HTML parsing* untuk mendapatkan dan mengumpulkan data informasi.
2. Objek penelitian berfokus untuk mendapatkan dan mengumpulkan informasi pada mesin pencari di *Google Scholar*.
3. Data yang didapatkan berupa informasi terkait artikel ilmiah yang akan dikumpulkan ke dalam suatu format file.

1.5 Metode Penelitian

Penelitian ini menggunakan teknik *Web Scraping* dengan metode *HTML parsing* untuk mendapatkan dan mengumpulkan informasi menggunakan bahasa pemrograman Python. Data yang diperoleh berupa informasi yang berkaitan dengan artikel ilmiah berdasarkan pencarian pada *Google Scholar*. Diharapkan hasil data yang diperoleh dapat membantu dalam pengumpulan data artikel ilmiah dengan hasil yang terstruktur dalam waktu yang singkat.

1.6 Jadwal Pelaksanaan

Pelaksanaan penelitian diperkirakan berlangsung selama lima bulan. Penelitian dimulai pada bulan Maret 2022 - Agustus 2022 secara bertahap. Berikut dijelaskan

jadwal pelaksanaan pada tabel dibawah inierisi jadwal pelaksanaan pengerjaan Tugas Akhir.

Tabel 1.1 Jadwal dan *Milestone*.

No.	Deskripsi Tahapan	Durasi	Tanggal Selesai	<i>Milestone</i>
1	Analisis web <i>Google Scholar</i>	2 minggu	5 Maret 2022	Bentuk program web <i>Google Scholar</i> dalam HTML
2	Belajar <i>Web Scraping</i> dengan bahasa pemrograman Python	7 minggu	23 April 2022	Program <i>Web Scraping</i> tahap awal
3	Implementasi metode HTML <i>parsing</i>	2 bulan	29 Juli 2022	Program <i>Web Scraping</i> selesai
4	Penyusunan laporan/buku TA	1 minggu	6 Agustus 2022	Buku TA selesai