

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Seiring berkembangnya zaman, kebutuhan akan informasi sangat tinggi. Berbagai upaya dilakukan untuk mendapat informasi dengan cepat dan tepat. Oleh karena itu pertumbuhan dari pengguna internet dalam pertukaran data sangat tinggi. Seiring perkembangan dari *World Wide Web*, banyak metode yang baru diperkenalkan untuk meningkatkan berbagai aspek mulai dari fasilitas komputer dan jaringan yang berdampak pada pengurangan biaya pada perangkat keras dan biaya lain yang berkaitan dengan pengembangan *website*. Karena itulah banyak pengguna dari berbagai kalangan seperti peneliti, pengusaha, akademisi saling menyebarkan informasi yang dimana hal ini bertujuan untuk menghubungkan informasi dengan cepat dan mudah. Namun hal ini menimbulkan pertanyaan baru, bagaimana menangani banyaknya data agar pengguna internet dapat mengakses informasi dengan mudah. Untuk menangani masalah tersebut para peneliti menemukan sebuah teknik yang dinamakan *web scraping* [1].

*Web scraping* merupakan teknik yang digunakan untuk mengekstrak data yang tidak terstruktur dari sebuah situs menjadi terstruktur. Metode dari *web scraping* juga berkembang seiring dengan berkembangnya *World Wide Web*. Sejak tahun 2000, *Document Object Model* (DOM) menjadi populer pada *Dynamic HTML*, yang kemudian mengubah teknik dari *HTML Parsing* menjadi *DOM Parsing*. Sebagai contoh lainnya adalah *Application Programming Interface (APIs)*. Beberapa metode *Web Scraping* lainnya di antaranya adalah *Manual Parsing*, *HTML Parsing*, *DOM Parsing*, *Xpath* dan *API*. [2].

Dalam melakukan *scraping* suatu *website*, hal yang menjadi perhatian utama adalah struktur data dari *website* tersebut dan mengidentifikasi apakah *website* tersebut dinamis atau statis. Dengan adanya bermacam-macam karakteristik *website*, tentu metode yang dipakai juga berbeda, sebagai contoh *website* saham atau *e-commerce* yang mana termasuk kedalam tipe *website* dinamis yang mana data akan berubah ubah secara *real-time*. Tentu hal itu

berbeda jika *website* tersebut merupakan *website* statis yang mana data diambil seperti artikel yang dipublikasikan secara *online*, informasi berbagai bidang atau bahkan profil perusahaan.

Penelitian tentang penggunaan metode *web scraping* pernah dilakukan oleh Ambre, Gaikwad, Pawar & Patil dengan judul penelitian “Web and Android for Comparison of E-Commerce Products” yang membahas mengenai mengambil data produk dari beberapa situs *e-commerce* berbeda dan membandingkannya untuk mencari harga yang diinginkan [3].

Dalam penelitian ini, penulis menganalisa beberapa metode yang digunakan untuk melakukan *scraping* data pada suatu *website* dan membandingkan metode-metode tersebut untuk mencari metode mana yang efektif dan efisien untuk struktur *website* tertentu.

## 1.2. Rumusan Masalah

Berdasarkan permasalahan di atas, dapat diambil rumusan masalah sebagai berikut:

1. Metode *Web Scraeping* mana yang lebih efektif dan cepat ?
2. Apakah ada perbedaan performa *Scraping* pada *website* statis dengan *website* dinamis ?

## 1.3. Tujuan

Berdasarkan permasalahan yang di atas, maka penelitian ini bertujuan untuk:

1. Menjelaskan dan membuktikan metode *Web Scraeping* mana yang lebih efektif dan cepat
2. Mejelaskan dan membuktikan perbedaan performa *Scraping* pada *website* statis dengan *website* dinamis.

## 1.4. Batasan Masalah

Berdasarkan dari rumusan masalah yang ada, penelitian ini memiliki batasan masalah, di antaranya:

1. Metode yang digunakan adalah *Regular Expression*, *Parsing DOM* and *XPath*

2. *Website* yang digunakan untuk uji coba performa adalah Wikipedia dan Kompas.com.
3. Dalam Penelitian ini berfokus pada uji performa dari metode *Web Scraping* dan menganalisa dan mengatasi masalah yang muncul ketika melakukan *Scraping* pada suatu *website*.

### 1.5. Metode Penelitian

Pada Tugas Akhir ini, penulis menerapkan beberapa metode untuk menyelesaikan masalah yaitu sebagai berikut:

1. Studi Literatur  
Langkah awal yang dilakukan adalah mempelajari tentang pengambilan data dan informasi yang didapat dari buku, jurnal, dan paper yang kemudian akan di review, dan informasi ini berkaitan dengan topik tugas akhir yang akan dilakukan penelitian ini.
2. Pengambilan Data  
Setelah dilakukan studi literatur dan mendapat sebuah data atau informasi maka dilakukanlah pengambilan data-data yang dapat menjadi parameter pendukung untuk perancangan sistem.
3. Perancangan Sistem  
Menentukan alur perancangan sistem dan mengimplementasikannya kedalam *environment* yang ada, dan kemudian akan dilakukan pengujian.
4. Pengujian Sistem  
Melakukan pengujian secara langsung dengan mengimplementasikan metode-metode untuk melakukan *web scraping*.
5. Analisis Sistem dan Metode  
Melakukan analisa dari hasil pengujian yang dilakukan sehingga dapat diketahui hasil yang diperoleh dari masing-masing metode dan dapat diketahui performa dari metode yang dilakukan apakah metode tersebut sudah sesuai dengan harapan.
6. Penyusunan Laporan Tugas Akhir  
Setelah didapat analisis data dan informasi dari metode yang diperlukan, selanjutnya dilakukan penyusunan laporan tugas akhir. Format laporan

tugas akhir akan mengikuti kaidah penulisan yang benar dan sesuai ketentuan-ketentuan pembuatan laporan tugas akhir yang telah ditentukan oleh institusi.

#### **1.6. Sistematika Penulisan**

Penulisan Laporan Tugas Akhir ini terdiri dari lima bab, dengan penjelasan sebagai berikut :

##### **BAB I PENDAHULUAN**

Pada bab ini dijelaskan mengenai latar belakang masalah yang mendasari adanya penelitian ini, identifikasi masalah, rumusan masalah, tujuan, metodologi, serta sistematika penulisan.

##### **BAB II TINJAUAN PUSTAKA**

Pada bab ini dijelaskan mengenai penjelasan teori yang digunakan untuk pengembangan penelitian ini. Yang berlandaskan dari pemikiran para Ahli

##### **BAB III PERANCANGAN SISTEM**

Pada bab ini dijelaskan tentang penjelasan mengenai rancangan sistem yang akan dibuat dalam penelitian yang dilakukan.

##### **BAB IV IMPLEMENTASI DAN PENGUJIAN**

Pada bab ini dijelaskan mengenai hasil dan implementasi sistem yang dibuat beserta pengujian apa yang dilakukan pada sistem.

##### **BAB V PENUTUP**

Pada bab ini dijelaskan mengenai kesimpulan dari hasil pengujian pada bab sebelumnya, serta saran untuk pengembangan aplikasinya.