

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Penggunaan mesin pencari atau *search engine* bagi user saat ini merupakan alat untuk mencari produk di internet berdasarkan kategori atau *keyword* tertentu. *Google* adalah salah satu *search engine* yang diminati, karena memiliki kelebihan dalam pencarian berdasarkan *web*, *image*, dan *news* dengan berbagai bahasa. Salah satu metode yang di gunakan dalam *search engine* adalah web scraping. Web Scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa *markup* seperti HTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain.

Bagi pengguna internet yang ingin mencari suatu produk menggunakan *search engine* dalam proses pencariannya. Hasil pencarian *search engine* yang menyeluruh kesemua jenis situs menyebabkan kurang efisien dan akan mengalami kesulitan jika ingin mendapatkan hasil pencarian dari situs jula beli saja. Pencarian dilakukan di satu persatu situs jual beli juga menyebabkan pemborosan kuota internet dan memakan waktu relatif lama.

Perlu adanya sebuah *search engine* yang dikhususkan hanya mencari informasi produk saja. *Serach engine* ini menggunakan HTML DOM (*Document Object Model*) yang digunakan sebagai alat untuk mengambil

informasi metadata yang mengacu pada *tag* HTML serta *regular expression(regex)* untuk menentukan pola yang mengawali dan mengakhiri suatu konten utama pada halaman *situs*. Kedua teknik tersebut sangat membantu dalam menemukan pola-pola kalimat. Sehingga percobaan terhadap semua kemungkinan pola pada kalimat tidak perlu dilakukan. Adapun kelebihan *search engine* ini adalah mempermudah pengguna internet dalam proses pencarian produk. Pencarian produk dilakukannya pada situs jual beli saja. Proses pencarian lebih cepat, mudah dan efisien, serta memberikan hasil pencarian yang sama dengan data pada situs yang diambil.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang masalah diatas maka dapat dirumuskan sebuah permasalahan yakni:

Bagaimana menerapkan teknik *web scraping* untuk mewedahi pencarian suatu produk di beberapa situs jual beli ke dalam satu bentuk *search engine* pencarian?

### **1.3. Tujuan Penelitian**

Adapun tujuan dari penulisan tugas akhir ini adalah membuat sebuah mesin pencari khusus produk yang hanya mengacu pada situs jual beli dengan menggunakan HTML DOM (*Document Object Model*) dan *regular expression(regex)* sebagai alat untuk melakukan proses *scraping* informasi produknya.

#### 1.4. Batasan Masalah

Untuk menghindari melebar nya pokok bahasan daripada penerapan teknik *web scraping*, yaitu :

a. Batasan situs jual beli. Situs jual beli yang akan di ambil datanya harus memilikisyarat :

1. Situs yang dipilih harus memiliki Memiliki fitur/kolom pencarian produk-produknya.
2. Situs yang dipilih harus memberikan hasil pencarian informasi produk tanpa di acak.
3. Pastikan *url* situs yang dipilih tidak memiliki *Secure Socket Layer*(SSL) karena jika sebuah situs memiliki SSL maka data atau semua informasi di dalamnya dalam kondisi teracak dan tidak bisa dilakukan pengambilan informasi data.
4. Pastikan Semua situs yang dipilih memiliki struktur *templateHTML*halaman hasil pencarian yang sama.

Berikut contoh ketiga situs yang memenuhi syarat di atas yang akan menjadi batasan objek penelitian adalah Jualo([www.jualo.co.id](http://www.jualo.co.id)), Bukalapak([www.bukalapak.com](http://www.bukalapak.com)), dan Bhinneka([www.bhinneka.com](http://www.bhinneka.com)).

b. Batasan teknis. Teknik *scraping* dilakukan pada dua halaman :

1. Pada halaman hasil url pencarian masing-masing situs jual beli dengan penambahan keyword pencarian.

2. Pada halaman detail *link* produk pencarian untuk mendapatkan informasi produk yang lebih lengkap.
3. Kategori informasi produk yang akan dicari terdiri dari : a) judul; b) gambar; c) harga; d) link; e) logo; f) *htmllink\_produk*; g) penjual; f) deskripsi; h) kondisi; i) lokasi;

### **1.5. Manfaat Penelitian**

Manfaat yang diharapkan dari penyusunan tugas akhir ini adalah Dengan adanya *search engine* pencari produk diharapkan pengguna internet yang ingin mencari produk hanya perlu singgah pada satu situs pencarian saja tanpa masuk ke semua situs jual beli.

### **1.6. Metodologi Penelitian**

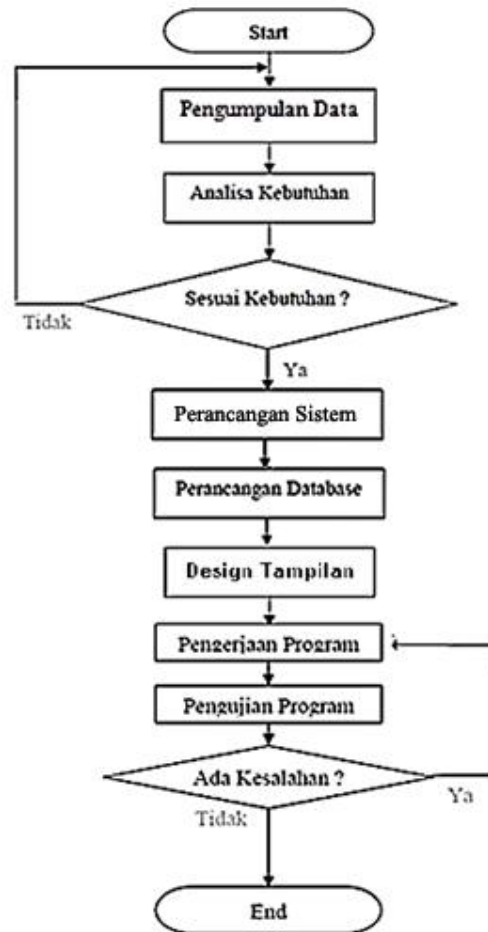
Berikut ini metode penelitian yang akan digunakan dalam penyusunan tugas akhir:

#### **1.6.1 Prosedur Penelitian**

Berikut prosedur penelitian yang digambarkan ke dalam flowchart prosedur penelitian :

1. Pengumpulan data, mencari pola url pencarian masing-masing situs jual beli. Menyisipkan *keyword* nama produk pada *url* pencarian dengan harapan memperoleh semua informasi produk sesuai *keyword*. Menentukan kategori informasi produk apa saja yang akan diambil.

Kategori produk seperti nama, harga, kondisi, nama pemilik, gambar dan sebagainya.



**Gambar 1.1** Flowchart Prosedur Penelitian

2. Analisa kebutuhan, dari pengumpulan data diperoleh *url* pencarian situs, kategori informasi produk beserta pola *tag* HTML pengapit. aktor-aktor apa saja yang terlibat pada aplikasi serta spesifikasi *use case* setiap aktor pada aplikasi. Jika data yang didapatkan tidak sesuai atau ada yang kurang maka akan kembali mengerjakan pengumpulan data sampai hasil yang didapatkan sesuai.

3. Perancangan Sistem, merancang atau mendesain suatu sistem yang baik, yang isinya adalah langkah-langkah operasi dalam proses pengolahan data dan prosedur untuk mendukung operasi sistem.
4. Perancangan Database, merancang database yang akan digunakan untuk menampung informasi-informasi terkait dengan *search engine*.
5. *Design interface*, yaitu tahapan memulai untuk membuat *design interface search engine*.
6. Pengerjaan Program, yaitu tahapan memulai mengerjakan program berdasarkan apa yang kita rancang.
7. Pengujian Program berupa pemeriksaan kembali antara *search engine* yang telah dibuat sesuai dengan apa yang direncanakan. Jika ditemukan *bug-bug* ataupun kesalahan maka kita akan memulai kembali pengerjaan program

Berdasarkan waktu dan prosedur penelitian yang telah ditentukan berikut ini rancangan jadwal pengerjaan tugas akhir:

**Tabel 1.1** Rancangan Jadwal Pengerjaan Tugas Akhir

Tahapan	Bln 1				Bln 2				Bln 3				Bln 4				Bln 5				Bln 6			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1																								
2																								
3																								
4																								
5																								
6																								



## BAB V : PENUTUP

Menentukan kesimpulan berdasarkan tujuan awal penyusunan tugas akhir dan beberapa saran yang dapat digunakan untuk kepentingan pengembangan sistem selanjutnya.