BAB I

PENDAHULUAN

1.1 Latar Belakang

Ilmu pengetahuan dalam bidang teknologi berkembang dengan sangat cepat, sehingga membuat orang semakin cepat dalam mengakses informasi. Informasi bisa didapatkan melalui koran, televisi ataupun internet. Situs pencarian seperti google dapat memudahkan pengguna dalam mencari informasi dimanapun dan kapanpun. Pada situs pencarian google, banyak informasi yang dapat di cari, seperti informasi berita dan jurnal, Banyaknya penyedia layanan artikel jurnal, membuat pengguna kesulitan saat ingin mencari artikel jurnal, karena harus mencari dengan cara mengunjungi satu-persatu situs jurnal. Google scholar adalah salah satu dari banyaknya situs pencarian jurnal. Google scholar dapat menampilkan semua artikel jurnal dari berbagai *website*, baik itu *website* terpercaya ataupun tidak. Jurnal akan menjadi sangat penting, baik itu sebagai media pembelajaran, ataupun sebagai referensi dalam melakukan penelitian. Menurut Efendi (2018) dalam situs *website*nya yang bernama *it-jurnal.com* menyebutkan bahwa google menjadi mesin pencari yang paling banyak digunakan yaitu dengan total pengunjung perbulan mencapai 30 milyar. Hal tersebut membuat situs pencarian google menjadi mesin pencari terpopuler pada tahun 2018, dan tentu saja akan terus menjadi mesin pencari terpopuler untuk beberapa tahun mendatang. Sistem kerja dari mesin pencari yaitu, melakukan pengumpulan informasi melalui program *bot* atau *web crawler* yang secara periodik menelusuri web.

Web crawler (Basori dkk, 2009) adalah sebuah program yang melintasi struktur *hypertext* dari web, dimulai dari sebuah alamat awal yang disebut *seed* dan secara rekursif mengunjungi alamat web di dalam halaman web, *web crawler* juga dikenal sebagai *web robot*, *spider*, *worm*, *walker* dan *wanderer*. Semua *search engine* besar menggunakan *crawler* yang mampu melintasi internet secara terus-menerus untuk menemukan dan mengambil halaman web sebanyak mungkin. Selain untuk *search engine, web crawler* juga digunakan untuk beberapa penelusuran khusus, seperti implementasi penelusuran alamat email. Adapun teknik dari *web crawler* yang disebut dengan "*web scraping*'.

Web scraping (Turland, 2010) adalah proses pengambilan suatu dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman *web* dalam bahasa *markup* seperti HTML, atau XHTML, dan menganalis dokumen tersebut untuk diambil data dari halaman tersebut untuk digunakan bagi kepentingan lain. *Web scraping* akan sangat tepat dalam penelitian ini, karena pengambilan data yang terstruktur dari teknik tersebut. Akan tetapi dalam menerapkan teknik *web scraping*, akan membutuhkan waktu lebih lama jika data yang dihasilnya dari *website* lebih banyak. Semakin banyak *hypertext* yang terdapat pada website tersebut akan semakin lama proses pengambilan data dari website tersebut, sehingga untuk mengambil data dalam jumlah besar diperlukan perangkat yang lebih baik.

Berdasarkan permasalahan dan teori yang dikemukakan, penulis akan menggunakan *web scraping* dalam pembuatan aplikasi untuk mengumpulkan artikel jurnal dari berbagai *website*, karena cara kerja dari metode *web scraping* yang tidak mengganggu *database* dari *website* yang akan diambil datanya, *web scraping* hanya akan mengambil dokumen *website* berdasarkan *element* pada *website* tersebut. Sehingga menggunakan *web scraping* akan lebih efisien untuk megumpulkan informasi artikel jurnal dari berbagai situs jurnal.

Berdasarkan hasil studi literatur yang telah penulis lakukan. Sebelum penelitian ini, telah banyak penelitian yang meneliti tentang penerapan teknik *web scraping*. Beberapa situs *website* yang di *scraping* untuk memudahkan pengguna dalam menghadapi masalah telah dipermudah dengan adanya beberapa penelitian tentang *web scraping*. Diantaranya adalah tentang mengumpulkan promo e-*commerce*, penjualan laptop dari beberapa situs jual beli, mengumpulkan data bencana alam, informasi pariwisata di daerah Yogyakarta, dan pencarian lowongan kerja yang diambil dari beberapa situs resmi penyedia lowongan kerja. Pada beberapa penelitian tentang teknik *web scraping* juga digabungkan dengan beberapa metode yaitu, *Breadt-First Search*, HTML DOM dan *Vector Space Model*. Setelah membaca beberapa paper yang mengangkat topik pembahasan tentang *web scraping* membuat penulis semakin tertarik dengan *web scraping* dan ingin mencoba untuk menerapkan secara langsung pada penelitian ini.

Dengan adanya aplikasi ini diharapkan akan lebih memudahkan pengguna dalam mencari artikel jurnal, pengguna tidak harus mencari artikel jurnal per situs untuk mendapatkan artikel jurnal yang dicari. Pengguna juga dapat melihat apakah judul dari artikel jurnal tersebut cocok dengan apa yang dibutuhkan, dan pengguna dapat langsung mengunduh artikel jurnal yang telah ditampilkan pada sistem.

1.2 Rumusan Masalah

Adapun rumusan masalah dari penelitian ini adalah, bagaimana penerapan teknik *web scraping* pada mesin pencari jurnal.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka tujuan yang ingin dicapai yaitu menerapkan teknik *web scraping* pada mesin pencari jurnal untuk mendapatkan info tentang judul jurnal, penulis, *issue* dan *link* unduh artikel jurnal.

1.4 Batasan Masalah

Batasan masalah dalam penerapan *web scraping* pada mesin pencari jurnal ini adalah :

1. Aplikasi dibuat hanya dalam bentuk *website*.
2. *Database* yang digunakan adalah *PhpMyAdmin*.
3. Objek yang akan di ambil datanya hanya website jurnal.
4. Output dari data yang diambil hanya, Judul Artikel, *Link Abstract* Penulis, *Issue*, dan *link download*.
5. Sumber jurnal untuk sementara hanya diambil dari situs portal garuda, portal IJoST(*Indonesian Journal of Sience and Technology*), dan portal jurnal tekno info.
6. Maksimal hanya mengambil 10 artikel jurnal dari setiap *website*.

1.5 Manfaat Penelitian

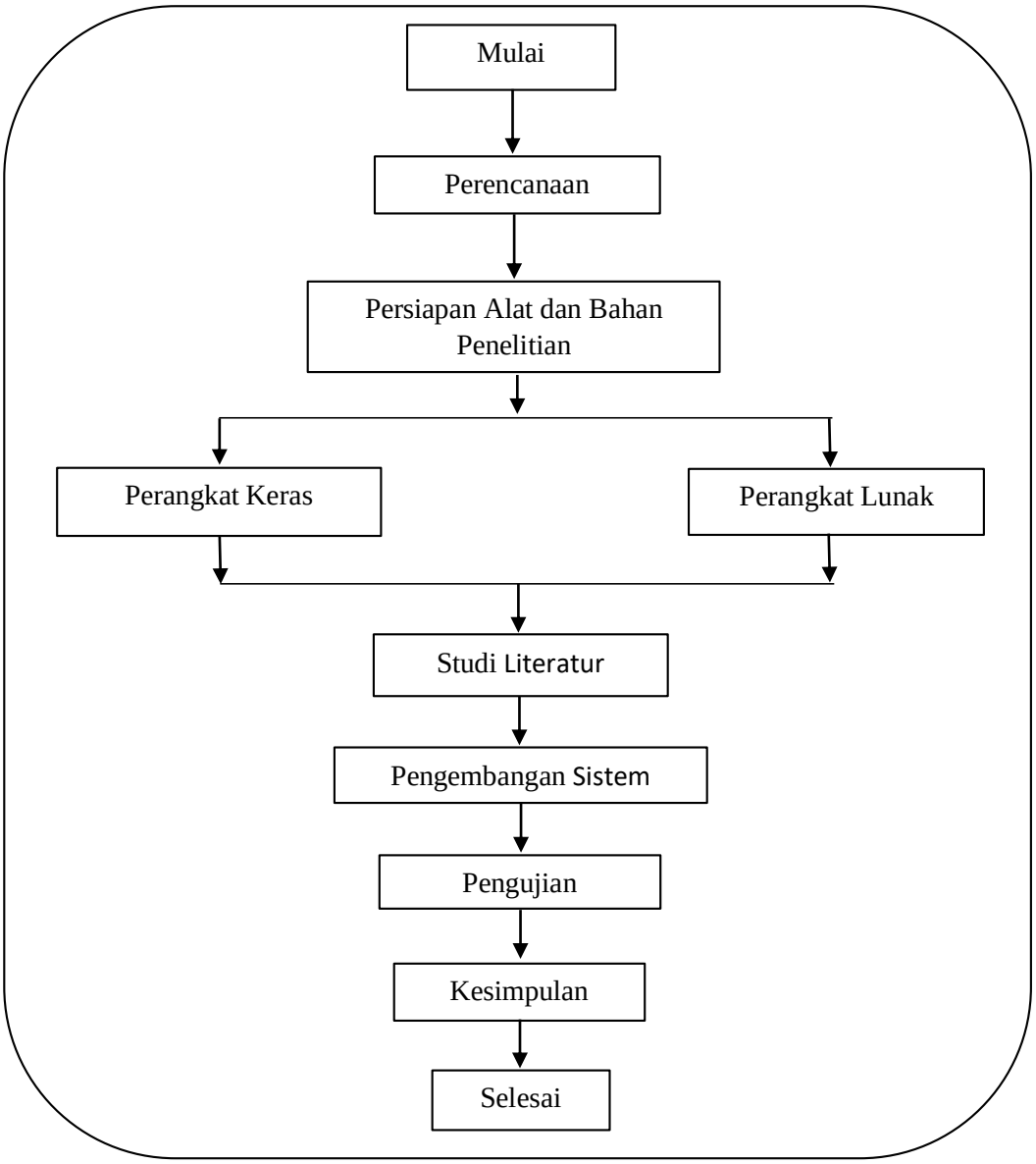
Penelitian ini diharapkan memberikan manfaat sebagai berikut :

1. Membantu pengguna internet untuk lebih mudah menemukan jurnal yang diinginkan tanpa harus mencari per situs jurnal.

2. Membantu penelitian tentang bagaimana menerapkan teknik web scraping sebagai sumber referensi yang dapat dikembangkan lebih lanjut.

1.6 Tahapan Penelitian

Tahapan penelitian mencakup langkah-langkah pelaksanaan penelitian dari awal sampai akhir. Langkah-langkah tersebut akan diuraikan secara rinci dan akan digambarkan dalam bentuk bagan sebagai berikut :



Gambar 1.1 Tahapan Penelitian

1.6.1 Perencanaan

Dalam tahap perencanaan penulis melakukan perencanaan tentang membangun aplikasi dengan menggunakan *web scraping*, dimulai dengan menggambarkan diagram alir dari sistem yang akan dibuat, untuk memudian dapat diterapkan kedalam bentuk *code*.

1.6.2 Persiapan Alat dan Bahan Penelitian

Tahapan pertama yang dilakukan dalam penelitian ini adalah mempersiapkan alat dan bahan penelitian berupa perangkat lunak dan perangkat keras yang digunakan untuk penlitian dan pengembangan perangkat lunak.

1.6.3 Studi Literatur

Tahapan selanjutnya adalah studi literatur dilakukan untuk mendapatkan dasar teori. Dasar teori didapatkan melalui jurnal, buku, dan internet. Penulis juga mempelajari tentang penelitian terdahulu yang membahas tentang *web scraping*. Untuk mengetahui apa saja yang dibahas dalam penelitian yang dilakukan tentang *web scraping*, dan untuk menjadi acuan pada penelitian yang akan penulis lakukan.

1.6.4 Pengembangan Sistem

Pada penelitian ini sistem yang dibuat akan menggunakan metode *Extreme Programming*. Metode *Extreme Programming* ini merupakan metode yang sangat *fleksibel* dan cocok dipakai untuk mengembangkan software dalam waktu singkat. Dalam metode ini terdapat empat tahapan dalam pengembangan sistem yaitu tahap *planning* (perencanaan), tahap desain, tahap *coding* dan tahap *testing*.

1.6.5 Pengujian

Setelah proses pengembangan sistem selesai, dan telah sesuai dengan apa yang dibutuhkan, maka selanjutnya akan dilakukan pengujian. Dalam pengujian hasil penerapan *web scraping*, proses pengujian yang penulis lakukan adalah dengan membandingkan hasil *scraping* dari sumber *website* yang dilakukan *scraping*, apakah data telah sesuai atau belum.

Pengujian selanjutnya adalah pengujian tentang respon sesorang atau sekelompok orang dari beberapa pertanyaan yang akan penulis buat dalam bentuk kuesioner, untuk mengetahui tanggapan dan respon dari responden. Dalam melakukan pengujian tersebut penulis menggunakan metode pengujian skala *likert*.

1.6.6 Analisa dan Penarikan Kesimpulan

Setelah pengembangan sistem, lalu dilakukan uji coba mengenai hasil dari *web scraping* yang telah diterapkan pada sistem yang dibuat. Lalu dilakukan analisa dan diambil kesimpulan dari penerapan *web scraping* tersebut terhadap sistem.

1.7 Sistematika Penulisan

Sistematika penulisan adalah penjelasan tentang apa saja yang ada pada setiap bagian penelitian mulai dari BAB I (satu) sampai dengan BAB V (lima). Sistematika penulisan pada penelitian ini akan disusun sebagai berikut :

1. BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang dari penelitian yang penulis lakukan, rumusan masalah, tujuan dari penelitian, batasan-batasan dari penelitian, manfaat yang didapatkan dari penelitian, dan tahapan yang dilakukan dalam melakukan penelitian.

2. BAB II LANDASAN TEORI

Pada bab ini menjelaskan tentang penelitian terdahulu yang membahas tentang *web scraping*, dan penjelasan tentang teori yang berhubungan dengan penelitian ini.

3. BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi tentang analisis dan perancangan sistem yang akan dibuat pada penelitian ini, seperti menganalisa data apa saja yang diperlukan, menganalisa *element* dari *website* jurnal, penjelasan dengan menggunakan diagram alir, mulai dari diagram alir proses *web scraping* dan diagram alir dari sistem yang akan dibuat.

Pada bab ini juga berisi tentang perancangan *prototype* sistem, seperti perancangan *interface*, dan perancangan data pada *database*.

4. BAB IV IMPLEMENTASI

Bab ini berisi tentang bagaimana melakukan implementasi mulai dari membuat *database*, desain tampilan, mengimplementasikan metode *web scraping* kedalam sistem berdasarkan dengan apa yang telah di analisa dan rencanakan pada penelitian ini. Dan pada bab ini juga ditulis tentang hasil dari pengujian yang akan penulis lakukan.

5. BAB V HASIL DAN PEMBAHASAN

Bab ini berisi tentang kesimpulan dari hasil implementasi yang telah dilakukan pada penelitian ini.