BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan teknologi informasi yang maju seperti sekarang ini membuat orang semakin cepat dalam mengakses informasi. informasi. Informasi bisa didapatkan lewati internet atau online. Informasi yang paling sering di akses adalah berita. Hal tersebut diimbangi dengan penyedia layanan situs berita online di Indonesia yang semakin banyak. Banyakk media cetak maupun televisi sekarang sudah mempunyai situs berita online sendiri.

Menurut (Putra, 2014), Situs berita online memudahkan pengguna dalam membaca berita dimanapun dan kapanpun. Kebanyakan informasi yang ada di internet banyak informasi yang tidak penting masuk di dalam berita tersebut, Sedangkan konten yang bagus dan unik dapat menarik banyak pengunjung website, itu merupakan kunci utama website atau situs online tersebut populer, Dengan banyaknya penyedia layanan situs berita online membuat pembaca harus berpindah situs berita untuk melihat berita yang bagus dan berbobot. Apalagi sekarang penyedia situs berita online tersebut mempunyai aplikasi mobile pembaca berita sendiri.

Menurut (Riyadi, 2014), Web Scraping merupakan teknik untuk mengambil informasi dalam suatu situs website secara otomatis. Fokus dari aplikasi Web Scraping adalah mengambil informasi dan mengekstrak informasi, Pengindeksan

website mempunyai hubungan dengan Web Scraping, tetapi Web Scraping fokus pada transformasi website tidak terstruktur menjadi format data terstruktur, Format data terstruktur tersebut dapat disimpan dan dapat dianalisa di database, Berita dari berbagai situs dapat diambil informasinya dengan teknik Web Scraping dan disimpan dalam database, Selain aplikasi pengambil informasi dari berbagai situs, dibutuhkan juga summarize berita untuk mendapatkan hasil summary beberapa berita menjadi satu.

Menurut (Favorisen Rosyking Lumbanraja, 2013), Semakin banyaknya koleksi dokumen teks, pencarian merupakan tantangan tersendiri. Banyak metode yang dikembangkan untuk proses pencarian, salah satu metode yang umum adalah dengan metode klasifikasi. Beberapa contoh teknik yang menggunakan metode klafisifikai antara lain, NaïveBayes, K-Nearest Neighbor, Decision Tree, dan Vector Space Model, Teknik Rocchio merupakan contoh lain yang mengimplementasikan metode klasifikasi untuk proses pencarian teks. Teknik ini menggunakan Vector Space Model untuk merepresentasikan setiap dokumen dalam korpus. Proses pertama yang dilakukan untuk mengembangan sistem dengan metode klasifikasi ini, yaitu tahap pra-proses.Pra-proses terdiri dari beberapa tahap, yaitu: parsering, pembersihan data, pemotongan kata berimbuhan, dan pembuatan inverted index dengan pembobot nilai itf.idf.

Permasalahan bahwa semua berita tidak terangkum dengan sesuai kriteria nya dan menyulitkan pembaca berita dalam mencari berita online sehingga orang sangat susah dalam mencari berita sesuai kriteria masing-masing, maka dari itu

membutuhkan aplikasi untuk dalam memilah berita di setiap situs yang dikunjungi.

Dari penjelasan di atas maka penelitian ini mencari situs berita untuk mengumpulkan berita dari berbagai situs dan meringkas berita. Maka dalam penulisan proposal ini diangkatlah sebuah judul yaitu **"Penerapan Web Scraping Berita Online".**

1.2 Perumusan Masalah

Dari uraian permasalahan diatas, maka penulis dapat merumuskan masalah yang ada untuk dijadikan titik tolak pembahasan dalam penulisan penelitian yaitu **"Bagaimana Penerapan *Web Scraping* Berita Online?".**

1.3 Batasan Masalah

Untuk menghindari terlalu luasnya ruang lingkup pembahasan dalam penelitian ini, maka penulis menetapkan batasan masalah sebagai berikut:

1. Mengetahui struktur halaman *website* yang akan di scrap atau diambil datanya.
2. Sumber berita untuk sementara hanya bisa diambil dari liputan6.com.
3. tools yang digunakan *webharvy*

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

Mendapatkan situs berita untuk mengumpulkan berita dari berbagai situs dan meringkas berita sesuai dengan kriteria pengguna.

1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

Membantu pembaca dalam memilih berita online yang sangat interaktir dan up to date.

1.5 Alat dan Bahan

Adapun alat dan bahan penelitian terdiri dari *hardware* dan *software* yang digunakan dalam Penerapan *Web Scraping* Untuk Mendapatkan Judul Berita Online ini adalah :

1. Perangkat keras (*hardware*) yang di gunakan dengan spesifikasi sebagai berikut :

   a. *Processor Intel I3*
   b. *RAM 2 GB*
   c. *Hardisk 500 GB*
   d. *Monitor SVGA Color*
   e. *CDRW Room 52 x*
   f. *Printer*
   g. *Keyboard*
   h. *Mouse*

2. Bahan yang digunakan berupa perangkat lunak (*software)* adalah sebagai berikut :

a. *Microsoft Windows 7* berfungsi sebagai operasi sistem
b. *MySQL* sebagai *database* untuk menyimpan data
c. *Microsoft Word 2007* sebagai aplikasi membuat laporan penelitian.

1.6 Metode Pengumpulan Data

Adapun metode pengumpulan data ini adalah sebagai berikut:

1. Observasi

Yaitu pengumpulan data dengan melakukan suatu pengamatan secara langsung pada link berita online yang akan menjadi objek penelitian.

2. Dokumentasi

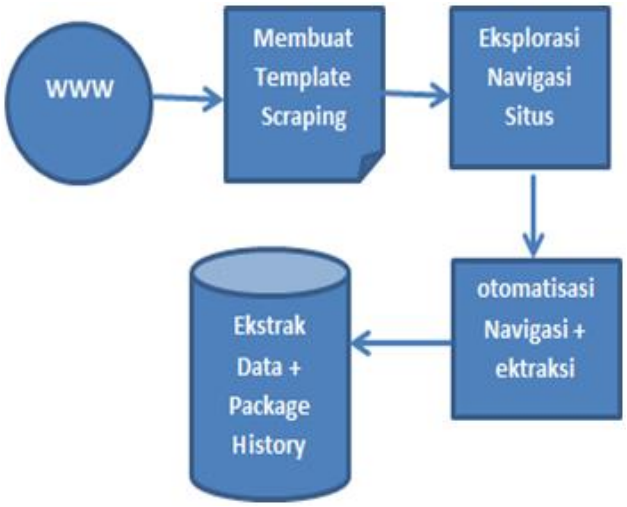
Yaitu metode yang digunakan untuk mengumpulkan dan mendapatkan sejumlah informasi yang berasal dari data-data. Data berita online yang yang meliputi liputan 6.

1.7 Proses *Web Scraping*

Web Scraping bukanlah data mining karena data mining adalah proses pengambilan informasi untuk memahami pola semantik atau tren dari sejumlah data yang besar (big data). Aplikasi *Web Scraping* atau *intelligent*, *automated*, or *autonomous* agent fokus pada cara memperoleh data melalui pengambilan data

Web Scraping memiliki sejumlah langkah, sebagai berikut :

a. Membuattemplate scraping: Proses ini melakukan observasi terhadap dokumen HTML website yang akan diambil informasinya atau dikenai scraping. Caranya adalah dengan melakukan tag HTML untuk mengapit informasi yang akan diambil,
b. Eksplorasi Navigasi Situs: Proses ini melakukanmenelusuri navigasi pada website yang akan diambil informasinya atau dikenai scrapinguntuk ditirukan pada aplikasi web scraper yang dibuat,
c. Mengotomatis Navigasi dan mengekstraksi informasi: Berdasarkan informasi yang didapat pada langkat 1 dan 2 di atas, aplikasi web scraper dibuat untuk mengotomatisasi pengambilan informasi dari website yang ditentukan, dan
d. Ekstraksi data dan menyimpan histori: Informasi yang didapat dari langkah 3 disimpan dalam tabel atau tabel- tabel database. Cara kerjanya lihat gambar.



Gambar 1.8 Ilustrasi Cara Kerja *Web Scraping*

Pada gambar di atas dapat dilihat bahwa yang pertama kali perlu dilakukan adalah denganmembuat template scraping. Proses tersebut dilakukan dengan cara mempelajari dokumen HTML dari website yang akan diambil informasinya untuk di tag HTML-nya. Tujuan nya adalah untuk mengambil informasi. Setelah itu, proses berikutnya adalah dengan mengeksplorasi navigasi situs yang dikenai scrapping. Tujuannya adalah mempelajari teknik navigasi pada website yang akan diambil informasinya untuk ditirukan pada aplikasi Web Scraping yang dibuat.

Selanjutnya, proses berikutnya adalah melakukan otomatisasi informasi yang didapat dari website yang telah ditentukan atau bias disebut juga sebagai proses ekstraksi informasi. Setelah informasi berhasil di ektraksi maka proses berikutnya adalah melakukan penyimpanan informasi ke dalam basis data (Esktraksi Data dan menyimpan histori).