

# 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 bisa didapatkan lewat 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. Banyak media cetak maupun televisi sekarang sudah mempunyai situs berita *online* sendiri.

Situs berita *online* memudahkan pengguna dalam membaca berita dimanapun dan kapanpun. Kebanyakan informasi yang ada di internet banyak yang bertele-tele dan banyak informasi yang tidak penting masuk di dalam berita tersebut (Putra, 2014). Sedangkan konten yang bagus dan unik dapat menarik banyak pengunjung *website*, itu merupakan kunci utama *website* atau situs *online* tersebut populer (Hardiono, 2010). Dengan banyaknya penyedia layanan situs berita *online* membuat pembaca harus berpindah-pindah situs berita untuk melihat berita yang bagus dan berbobot. Apalagi sekarang penyedia situs berita *online* tersebut mempunyai aplikasi *mobile* pembaca berita sendiri-sendiri. Dengan keadaan tersebut membuat pembaca berita harus memasang aplikasi lebih dari satu kalau ingin membaca berita dari berbagai sumber. Saat ini dibutuhkan aplikasi khusus untuk menggabungkan beberapa situs tersebut menjadi satu. Teknik *web scraping* tersebut yang cocok

digunakan untuk mengambil berita dari berbagai situs tersebut.

*Web scraping* merupakan teknik untuk mengambil informasi dalam suatu situs *website* secara otomatis (Riyadi, 2013). 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 (Zaira, 2010). 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.

*Summarize* berita berfungsi untuk meringkas beberapa berita menjadi satu paragraf. Paragraf yang didapatkan merupakan inti dari beberapa berita. Ringkasan berita bisa dari sumber yang sama dan bisa dari sumber yang berbeda. Peringkasan Berita dari berbagai sumber merupakan proses penyaringan informasi penting dalam satu set dokumen untuk menghasilkan informasi yang relevan bagi pengguna (Evan, et al., 2014). Pengguna akan lebih cepat mengetahui inti berita yang dicari dengan fitur *summarize* berita ini. *Summarize* berita pada penelitian ini menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Peringkasan berita ini akan mengambil data yang dikumpulkan oleh *web scraper* di *database*.

Aplikasi yang dibangun memanfaatkan teknologi *web scraping* untuk mengambil data pada berbagai situs berita. Data berita yang diambil tersebut langsung masuk ke dalam *database* untuk disimpan. Data berita dalam *database* dapat ditampilkan pada aplikasi. Selain ditampilkan, data berita dapat diolah untuk di ringkas menjadi lebih mudah mengetahui inti yang dicari. Aplikasi untuk meringkas berita memanfaatkan metode TF-IDF. TF-IDF digunakan karena dapat memberikan nilai pada kalimat yang sebagai inti pada sebuah berita. Aplikasi untuk meringkas dan menampilkan berita dari berbagai sumber berjalan pada *mobile device*. Aplikasi *mobile* tersebut digunakan oleh pengguna secara umum. Pada sisi *administrator* akan mengatur *scraping website* yang berisi pengaturan URL (*Uniform Resource Locator*) dan *rules*.

Dengan adanya aplikasi ini memudahkan pengguna dalam membaca berita. Pengguna tidak harus melihat berita per situs atau mengunduh semua aplikasi berita, hanya dengan satu aplikasi dapat melihat semua berita dari berbagai sumber. Selain dapat melihat berita dari berbagai sumber, ada fitur peringkas berita yang akan memudahkan pengguna dalam mengetahui inti berita yang dicari.

## **1.2 Rumusan Masalah**

Berdasarkan dari latar belakang di atas, maka dapat di rumuskan permasalahannya sebagai berikut:

Bagaimana membangun aplikasi *web scraper* situs berita untuk mengumpulkan berita dari berbagai situs dan meringkas berita sesuai kriteria pengguna?

## **1.3 Batasan Masalah**

Batasan masalah dalam aplikasi *web scraper* situs berita ini adalah :

1. Aplikasi terbagi menjadi 2 yaitu aplikasi untuk admin berbasis *website* dan aplikasi *user* pembaca berita berbasis *mobile* android.
2. Admin harus mengetahui struktur halaman *website* yang akan di *scrap* atau diambil datanya.
3. Sumber berita untuk sementara hanya bisa diambil dari *kompas.com*, *tempo.co*, *okezone.com*, *sindonews.com*, *liputan6.com*, dan *detik.com*.

## **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah di atas maka ada beberapa tujuan yang ingin dicapai yakni:

Membangun aplikasi *web scraper* situs berita untuk mengumpulkan berita dari berbagai situs dan meringkas berita sesuai dengan kriteria pengguna.

## 1.5 Metodologi Penelitian

Metode yang digunakan dalam pembangunan aplikasi ini yakni :

### 1. Kepustakaan

Metode ini dilakukan dengan cara mengumpulkan dan mempelajari berbagai sumber referensi yang ada, seperti buku-buku dan artikel-artikel di internet yang mendukung dalam perancangan dan implementasi aplikasi yang dibuat.

### 2. Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak mencakup proses spesifikasi kebutuhan perangkat lunak, seperti antarmuka dengan pengguna (*user interface*) maupun kinerja (*performance*) perangkat lunak pada berbagai fungsi yang dirancang untuk dapat dilaksanakan oleh sistem. Dalam metode ini ada 4 sub metode, yaitu:

#### 2.1 Analisis

Analisis kebutuhan merupakan bagian yang tidak bisa dipisahkan dari pengembangan perangkat lunak, dimana perangkat lunak yang akan dibuat nantinya akan disesuaikan dengan spesifikasi dan sistem-sistem pendukung lainnya, baik itu perangkat keras maupun perangkat lunak. Hasilnya berupa spesifikasi kebutuhan perangkat lunak (SKPL).

#### 2.2 Perancangan

Perancangan dilakukan untuk mendiskripsikan bagaimana perangkat lunak akan dibangun. Perancangan yang dilakukan mulai dari perancangan arsitektur, perancangan antarmuka,

dan perancangan basis data. Hasilnya berupa Deskripsi Perancangan Perangkat Lunak (DPPL).

### 2.3 Pengkodean

Pengkodean adalah menyalin alur data dan alur sistem kedalam bahasa pemrograman. Dalam penelitian ini bahasa pemrograman yang digunakan adalah PHP (*PHP:Hypertext Preprocessor*) framework CodeIgniter, HTML5 (*HyperText Markup Language 5*), CSS3 (*Casading Style Sheet 3*), dan Javascript. Pengkodean yang dilakukan pertama adalah pengkodean alur data, kemudian dilanjutkan dengan pengkodean alur sistem.

### 2.4 Pengujian

Pengujian akan dilakukan dengan cara menguji fungsionalitas program yang dibangun. Pengujian dibagi menjadi 2 macam yaitu pengujian oleh ahli yaitu programmer dan pengujian oleh user. Hasil pengujian berupa Perancangan, Deskripsi dan Hasil Uji Perangkat Lunak (PDHUPL).

## 1.6 Sistematika Penulisan

Laporan ini ditulis dengan sistematika sebagai berikut :

### **BAB 1 : Pendahuluan**

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan, metodologi penelitian, dan sistematika penulisan laporan.

## **BAB 2 : Tinjauan Pustaka**

Bab ini berisi penjelasan mengenai penelitian yang pernah dilakukan sebelumnya yang berkaitan dengan topik yang dibahas, dan penjelasan mengenai perbandingan antara penelitian yang sudah dilakukan sebelumnya dengan penelitian yang akan dilakukan.

## **BAB 3 : Landasan Teori**

Berisi penjelasan mengenai dasar teori yang berkaitan dengan permasalahan yang dibahas.

## **BAB 4 : Analisis dan Perancangan Perangkat Lunak**

Bab ini berisi penjelasan mengenai analisis permasalahan yang akan diatasi serta membahas mengenai perancangan perangkat lunak yang dibuat.

## **BAB 5 : Implementasi dan Pengujian Perangkat Lunak**

Bab ini berisi penjelasan mengenai implementasi perangkat lunak yang dibuat dan gambaran umum sistem.

## **BAB 6 : Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari pembahasan secara keseluruhan beserta saran-saran yang bermanfaat untuk pengembangan lebih lanjut.

## **DAFTAR PUSTAKA**

## **LAMPIRAN**