

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar belakang**

Perkembangan informasi teks digital telah tumbuh sangat cepat. Saat ini diperkirakan 80% teks digital dalam bentuk tidak terstruktur. Tingginya volume dokumen teks ini dipicu oleh aktivitas dari berbagai sumber berita dan aktivitas akademis dari kegiatan riset, konferensi dan pertemuan ilmiah yang makin meningkat (Hamzah, 2012).

Dengan adanya internet, informasi dapat dengan mudah disebarluaskan dan diakses oleh banyak orang. Banyaknya informasi yang beredar tentu membuat kebutuhan akan informasi yang relevan semakin meningkat (Maarif, 2015).

Peringkasan sebuah artikel merupakan sebuah cara pengambilan informasi dari sebuah dokumen teks dengan mengambil sebagian teks dalam dokumen yang dianggap mengandung informasi paling penting. Selain itu, dengan adanya ringkasan, manusia dapat dengan mudah dan lebih cepat mengerti dan memahami isi sebuah dokumen tanpa harus membaca keseluruhan dokumen yang pasti memerlukan waktu yang lebih lama. Maka dari itu diperlukan sebuah sistem yang mampu meringkas sebuah dokumen teks secara otomatis (Maharani & Sanjaya, n.d.).

Perkembangan teknologi internet berdampak bertambahnya jumlah situs berita berbahasa

Indonesia dan menciptakan ledakan informasi. Hal tersebut menuntut semua informasi bisa diakses dengan cepat dan tidak harus membutuhkan banyak waktu dalam sebuah headline berita (Mustaqhfiri, et al., 2011). Peringkasan multi-dokumen dibutuhkan agar pencarian informasi dapat berjalan lebih efektif dan efisien (Akbar, et al., 2015).

Peringkasan yang baik terhadap dokumen-dokumen dengan topik yang seragam dapat membantu pembaca dalam memperoleh informasi secara cepat. Ringkasan yang baik merupakan ringkasan dengan cakupan pembahasan yang luas dan dengan tingkat keberagaman serta keterhubungan antarkalimat yang tinggi (Umam, et al., 2015).

Informasi dalam bentuk teks berita telah menjadi salah satu komoditas yang paling penting dalam era informasi ini. Ada banyak berita yang dihasilkan sehari-hari, tetapi berita-berita ini sering memberikan konten kontekstual yang sama dengan narasi berbeda. Oleh karena itu, diperlukan metode untuk mengumpulkan informasi ini ke dalam ringkasan sederhana (Lukmana, et al., 2014) .

Studi pustaka penelitian di bidang klasterisasi dan klasifikasi dokumen teks berbahasa Indonesia menunjukkan bahwa penelitian bidang pemrosesan dokumen telah dimulai pada tahun 2000. Terdapat berbagai metode data mining untuk melakukan pengelompokan dokumen digunakan seperti *single pass filtering*, *Naïve Bayes*, Hirarki dan metode lainnya (Raharjo & Winarko, 2014).

Sebelumnya sudah dibangun aplikasi peringkas berita yang menjadi sub sistem dari KUBERIN namun pada saran penulis berharap aplikasi KUBERIN dapat dikembangkan lebih lanjut untuk bagian summarize berita agar dapat menghasilkan hasil ringkasan yang baik dan dimengerti oleh pengguna, misalnya dengan mengidentifikasi kemiripan dokumen terlebih dahulu sebelum diringkas (Setiawan, 2015).

Aplikasi peringkas dokumen juga sudah pernah dibangun sebelumnya, namun aplikasi ini memiliki kekurangan pengambilan kalimat yang akan digunakan sebagai ringkasan akhir sebaiknya diambil dari masing-masing dokumen sumber sehingga inti dari masing-masing dokumen terlihat pada ringkasan akhir, dan Penambahan metode untuk menggabungkan kalimat agar kesinambungan antar kalimat dapat menjadi lebih baik lagi (Evan, 2014).

Penelitian ini akan mencoba melakukan peringkasan banyak dokumen dengan terlebih dahulu melakukan klusterisasi. Hasil peringkasan akan dibandingkan dengan metode yang dikerjakan sebelumnya yaitu dengan sentence scoring. Dari penelitian ini adalah mengetahui mana yang lebih baik.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan maka diperoleh beberapa masalah sebagai berikut:

1. Bagaimana membangun aplikasi *web* peringkasan dokumen dengan Maximum Marginal Relevance?

2. Bagaimana membandingkan peringkasan banyak dokumen metode Sentence Scoring dengan Maximum Marginal Relevance dengan proses klustering terlebih dahulu?

### **1.3 Batasan Masalah**

Aplikasi web yang diajukan peneliti memiliki batasan masalah yaitu:

1. Dokumen yang dikelompokkan dan diringkas hanya berbahasa Indonesia
2. *Framework* yang digunakan merupakan Code Igniter

### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Membangun aplikasi web peringkasan dokumen dengan Maximum Marginal Relevance.
2. Membandingkan peringkasan banyak dokumen metode Sentence Scoring dengan Maximum Marginal Relevance dengan proses klustering terlebih dahulu.

### **1.5 Metodologi Penelitian**

Metode yang digunakan dalam pembangunan aplikasi web ini adalah:

#### a. Studi Pustaka

Metode ini dilakukan dengan cara mencari, mengumpulkan, dan mempelajari berbagai sumber referensi yang ada, seperti buku-buku dan artikel-artikel yang berkaitan dengan aplikasi yang dikembangkan. Referensi digunakan untuk membantu mempertegas teori-teori yang ada sehingga memperoleh data yang sesungguhnya.

b. Perancangan Sistem

Metode perancangan ini merupakan metode untuk merancang aplikasi yang terdiri dari arsitektural perangkat lunak, antarmuka, data dan prosedural. Hasil dari metode perancangan ini berupa Deskripsi Perancangan Perangkat Lunak (DPPL).

c. Implementasi Sistem

Metode implementasi ini digunakan untuk mengimplementasikan proses pembangunan aplikasi dari hasil rancangan yang telah dibuat. Proses penerjemahan dari desain yang telah dibuat ke bahasa pemrograman.

d. Pengkodean

Pengkodean adalah menyalin alur data dan alur system 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 system.

e. Dokumentasi

Metode dokumentasi ini merupakan metode yang digunakan untuk mencatat data yang telah terkumpul. Data tersebut berguna untuk pembangunan perangkat lunak ke dalam bentuk dokumen.

## **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**