

BAB I PENDAHULUAN

1.1 Latar Belakang

Berita adalah kumpulan informasi yang menyampaikan peristiwa yang sedang terjadi. Berita mengandung fakta atau opini yang membuat orang tertarik untuk mengetahuinya. Berita dapat diperoleh dari media seperti televisi, koran, surat kabar, internet, dan lain-lain. Saat ini, media yang sering digunakan untuk memperoleh berita adalah internet. Dengan internet, semua berita dapat diperoleh dengan mudah, cepat, dan *up-to-date*. Karena semakin banyak masyarakat yang mengakses berita melalui internet, sehingga jumlah portal berita online meningkat. Portal berita online menyajikan berita mulai dari skala lokal, nasional bahkan internasional. Dalam memperoleh informasi di berita, maka diperlukan sebuah keterampilan membaca. Membaca adalah suatu proses yang dilakukan oleh pembaca untuk memperoleh pesan atau informasi yang hendak disampaikan penulis melalui bahasa tulis [1]. Semakin banyak berita yang disuguhkan, maka lebih banyak waktu yang digunakan untuk mendapatkan informasinya. Oleh karena itu, di samping keterampilan membaca, diperlukan peringkasan kumpulan berita secara otomatis agar lebih efisien dan efektif dalam memperoleh informasi dari berita.

Peringkasan teks otomatis adalah pembuatan bentuk yang lebih singkat dari suatu teks dengan memanfaatkan aplikasi yang dijalankan dan dioperasikan oleh computer [2]. Pada peringkasan teks terdapat dua pendekatan yaitu ekstraksi (*shallower approaches*) dan abstraksi (*deeper approaches*) [3]. Terdapat kategori peringkasan teks dari sebuah *automatic text summarization* salah satunya adalah *query-focused summarization* [4]. *Query-focused summarization* adalah peringkasan suatu teks yang diberikan *query* untuk menghasilkan ringkasan teks tersebut, sehingga peringkasan teks ini relevan terhadap *query* yang diberikan [5]. Terdapat permasalahan yang dialami pada peringkasan teks tersebut yaitu menimbulkan ambiguitas dan redundansi data. Hal tersebut tentu memengaruhi kualitas dari hasil peringkasan teks. Sehingga dalam mengatasi masalah ini

diperlukan sebuah metode yang dapat menghubungkan antara *query* dengan teks serta dapat mengurangi ambiguitas dan redundansi data.

Beberapa metode peringkasan teks otomatis telah banyak dilakukan dengan berbagai metode dan diterapkan di berbagai bahasa. Di antaranya adalah *word frequency* [6], *sentence position* [7], *lexical chain* [8], dan *maximum marginal relevance* [9]. Khusus untuk teks berbahasa Indonesia, metode dan algoritma yang digunakan antara lain algoritma genetika [10], metode *relevance measure*, *latent semantic analysis* [11] dan *maximum marginal relevance* [2]. Pada penelitian kali ini, peneliti akan melakukan peringkasan secara ekstraktif dengan mengidentifikasi bagian informasi penting berdasarkan statistika penghitungan frekuensi kemunculan kata dengan metode TD-IDF. Kemudian setelah menghitung frekuensi kemunculan kata, penelitian akan dikembangkan menggunakan metode pendekatan ekstraksi menggunakan algoritma *Maximum Marginal Relevance*.

Maximum Marginal Relevance (MMR) merupakan salah satu metode *extractive summarization* yang digunakan untuk meringkas *single document* atau *multi document* . [12]. MMR meringkas dokumen dengan menghitung kesamaan (*similarity*) antara kalimat dengan kalimat dan antara kalimat dengan *query* [2]. MMR adalah salah satu teknik peringkasan yang ditujukan untuk mengambil sebuah informasi yang relevan dan tidak mengandung redundansi (perulangan) [13]. Peringkasan MMR mengombinasikan *cosine similarity* antar bagian kalimat. Dalam memilih sebuah *query* perlu dipertimbangkan aspek kerelevanan antara dokumen dengan *query* untuk mendapatkan hasil ringkasan yang baik [5]. Ide dasar dari MMR ini yaitu memberikan penambahan nilai bagi kalimat yang relevan dan memberikan pengurangan nilai redundansi informasi antara kalimat tersebut dengan kalimat lain yang telah terpilih [9]. Sebuah kalimat bisa dikatakan mempunyai *marginal relevance* yang tinggi jika kalimat tersebut relevan terhadap isi kalimat dan mempunyai kesamaan bobot *term* maksimum dibandingkan dengan *query* [14].

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dikemukakan, maka bisa dirumuskan beberapa rumusan masalah berikut ini :

- a. Bagaimana penerapan peringkasan teks otomatis menggunakan algoritma *Maximum Marginal Relevance*?
- b. Bagaimana tingkat akurasi peringkasan teks yang dihasilkan menggunakan *Maximum Marginal Relevance*?

1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut :

- a. Peringkasan teks otomatis menggunakan metode *Maximum Marginal Relevance*.
- b. Penelitian ini menggunakan pembobotan berdasarkan jumlah kata dalam sebuah dokumen (*term frequency*) dan jumlah kemunculan kata yang muncul pada seluruh dokumen yang ada (*inverse document frequency*).

1.4 Tujuan Penelitian

Adapun tujuan peneliti dalam melakukan penelitian adalah sebagai berikut:

- a. Mengetahui penerapan metode *Maximum Marginal Relevance* untuk meringkas teks secara otomatis.
- b. Menunjukkan tingkat akurasi metode *Maximum Marginal Relevance* dalam melakukan peringkasan teks secara otomatis.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini antara lain :

- a. Mempermudah pembaca dalam menghemat waktu pada saat membaca dan memahami berita.
- b. Dapat mengetahui nilai akurasi yang didapat dari penggunaan metode *Maximum Marginal Relevance* dalam peringkasan teks otomatis.
- c. Penelitian ini diharapkan dapat menjadi referensi bagi penulis atau pengembang untuk penelitian di masa yang akan datang.
- d. Menambah wawasan serta ilmu pengetahuan dalam pengembangan

sistem yang lebih lanjut.

1.6 Sistematika Penulisan

Sesuai dengan petunjuk penulisan laporan skripsi yang berlaku di Universitas Amikom Yogyakarta, sistematika penulisan laporan ini adalah sebagai berikut yang berisi paparan garis besar isi masing - masing bab.

BAB I PENDAHULUAN

Bab pendahuluan berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab tinjauan Pustaka berisi teori atau konsep yang digunakan sebagai landasan berpikir dalam penelitian ini. Tinjauan pustakan diperoleh dari buku, jurnal, dan beberapa penelitian sebelumnya yang terkait dengan topik *Automatic Text Summarization*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan alur tahap penelitian, alat berupa algoritma, serta perangkat dan bahan berupa data yang akan digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan berisi hasil penelitian secara menyeluruh serta pembahasan pada penelitian yang dikaji.

BAB V KESIMPULAN DAN SARAN

Bab kesimpulan dan saran berisi mengenai kesimpulan yang didapat dari penelitian di bab-bab sebelumnya, serta saran yang diharapkan dapat berguna untuk pengembangan penelitian di masa mendatang.