

1. Pendahuluan

1.1 Latar Belakang

Dengan semakin berkembangnya ilmu pengetahuan dan teknologi, membuat jumlah informasi menjadi begitu banyak, salah satunya informasi berupa teks terutama pada media elektronik. Oleh karena itu, dibutuhkan informasi yang singkat dan padat yang merepresentasikan isi dokumen. Sehingga dikembangkan sebuah sistem yang mampu meringkas dokumen yaitu, peringkasan teks otomatis (*automatic text summarization*).

Ikhtisar atau yang biasa disebut dengan ringkasan adalah suatu informasi penting dari suatu sumber atau sumber informasi ganda menurut kebutuhan-kebutuhan tertentu. Dengan ringkasan, kita dapat membuat keputusan-keputusan efektif dan mendapat informasi bermanfaat dalam waktu lebih singkat. Peringkasan sudah dimulai sejak tahun 1950. Edmundson menyajikan suatu survei dari metode-metode yang ada untuk ikhtisar yang otomatis (*automatic summarization*).

Banyak metode maupun pendekatan yang digunakan untuk melakukan peringkasan teks otomatis. Terdapat dua buah pendekatan dilihat dari teknik pengambilan ringkasan, yaitu ekstraksi dan abstraksi. Di mana ekstraksi merupakan teknik menyeleksi materi dari sumber yang berupa teks sedangkan abstraksi merupakan teknik meringkas teks dengan cara mereformulasikan kembali versi aslinya.

Tugas Akhir ini menggunakan pendekatan ekstraksi dalam melakukan peringkasan. Ringkasan dari suatu teks diekstraksi dengan melakukan pemilihan dari bagian dokumen yang penting untuk menghasilkan hasil yang lebih singkat. Ringkasan manusia sering kali dilakukan dengan cara meng-*cut* dan *paste* dari suatu dokumen untuk menghasilkan ringkasan. Kita dapat belajar jenis operasi yang biasanya dilaksanakan secara manual yang dilakukan manusia untuk mengekstraksi dan meng-edit kalimat-kalimat kemudian mengembangkan program yang dapat bekerja secara otomatis (*automatic programs*) untuk menirukan operasi tersebut. Granularitas-granularitas dari ekstraksi berupa frase-frase (2 atau 3 kata-kata) seperti pada kalimat. Pendekatan ekstraksi mungkin punya permasalahan pada koherensinya. Salah satu pendekatan ekstraksi yaitu dengan menggunakan metode pembobotan TF-IDF dan metode LSA.

Beberapa metode pembobotan kata secara umum yang biasa digunakan untuk teks yaitu *term frequency* (TF) dan *inverse document frequency* (IDF). Pembobotan TF IDF (*relevance measure*) memperhitungkan kemunculan term tidak hanya pada kalimat yang memiliki term tersebut, tetapi menimbang kemunculan term dalam keseluruhan dokumen.

Latent Semantic Analysis (LSA) yang diilhami dari pengindeks laten semantik dan menerapkan *Singular Value Decomposition* (SVD) ke dokumen sentence matrix [7]. *Singular Value Decomposition* (SVD) adalah suatu alat mathematical sangat tangguh untuk menemukan dimensi-dimensi pokok orthogonal data multidimensional. SVD mempunyai aplikasi-aplikasi di dalam banyak bidang dan dikenal oleh nama-nama yang berbeda, *Latent Semantic Analysis* (LSA) di dalam pengolahan teks. SVD dalam pengolahan teks diberi nama LSA karena SVD

berlaku untuk *document-word matrices*, kelompok-kelompok dokumen yang bersifat secara semantis berhubungan dengan satu sama lain.

Tugas Akhir ini dilakukan dua metode text summarization untuk membuat *generic text summaries* dengan peringkat (*ranking*) dan mengekstraksi kalimat-kalimat dari dokumen asli. Metode yang pertama menggunakan metode *relevance measure* untuk me-rank keterkaitan kalimat dan metode yang kedua gunakan *latent semantic analysis* yang dapat mengidentifikasi kalimat-kalimat penting secara semantis, untuk membuat ikhtisar.

1.2 Perumusan Masalah

Dalam pengerjaan Tugas Akhir ini, ada beberapa rumusan masalah yang coba untuk diselesaikan, yaitu sebagai berikut :

- a) Bagaimana mengimplementasikan ringkasan teks yang *generic* menggunakan *relevance measure* and *Latent Semantic Analysis*?
- b) Bagaimana tingkat akurasi hasil ringkasan tersebut?

Dengan mempertimbangkan kompleksitas yang mungkin ada pada suatu ikhtisar, maka pada Tugas Akhir ini terdapat beberapa batasan masalah, yaitu :

- a) Membuat ikhtisar dengan menggunakan metode *Relevance Measure* dan metode Latent Semantic Analysis (untuk mengidentifikasi *important sentences*)
- b) Menggunakan dokumen tunggal teks berita berbahasa Indonesia.
- c) Peringkasan teks dilakukan secara *offline*.
- d) Ringkasan berupa hasil ekstraksi kalimat-kalimat dari dokumen asli.
- e) Tidak melakukan *stemming* terhadap teks masukan.
- f) Hasil ringkasan yang diperoleh dibandingkan dengan hasil ringkasan manual.

1.3 Tujuan

Tujuan yang ingin dicapai dari Tugas Akhir ini yaitu :

- a) Menganalisis dan mengimplementasikan *Relevance Measure* dan *Latent Semantic Analysis* untuk membuat ringkasan otomatis.
- b) Menganalisis dan membandingkan performansi hasil ringkasan berdasarkan dua metode tersebut berupa akurasi berdasarkan *precision*, *recall*, dan *f-measure* dengan menggunakan Rouge.

1.4 Metodologi Penyelesaian Masalah

Dalam menyelesaikan Tugas Akhir ini, ada beberapa tahapan metode akan dilakukan, yaitu :

- a) Studi Literatur
Studi literatur bertujuan untuk meningkatkan pemahaman terhadap metode-metode yang akan digunakan dalam Tugas Akhir, yaitu metode *Relevance measure* dan *LSA*, dengan cara mencari referensi yang berkaitan dengan metode-metode tersebut dan kemudian mendalami materinya.
- b) Analisis dan Perancangan Perangkat Lunak

Pada tahap ini akan dilakukan analisis terhadap perangkat lunak yang diperlukan dalam membangun implementasi dari *Summarization*, baik berupa bahasa pemrograman yang akan digunakan sampai pada algoritma dan struktur data yang digunakan.

c) Implementasi

Hasil perancangan pada tahap sebelumnya akan menjadi dasar dalam tahap implementasi ini. Pada tahap ini akan dilakukan pengkodean berdasarkan hasil rancangan di atas.

d) Testing

Di tahap ini, akan dilakukan pengujian terhadap hasil implementasi guna menemukan dan menghilangkan *error/bug* yang mungkin masih ada.

e) Analisis Hasil Implementasi

Analisis hasil implementasi dilakukan dengan mengukur akurasi atau ketepatan kerja dari program dalam mengikhtisarkan suatu dokumen

f) Pembuatan Laporan Tugas Akhir.

Pada tahap akhir, dilakukan pembuatan dokumentasi yang berupa laporan Tugas Akhir.