

1. Pendahuluan

1.1 Latar Belakang Masalah

Semakin berkembangnya dunia komputerisasi sehingga menyebabkan semakin cepat pula informasi yang didapat. Suatu informasi merupakan hal yang sangat penting dan berguna oleh setiap orang, salah satu bentuk informasi bisa di dapatkan dari media online seperti artikel-artikel berita online. Sangat banyak berita online memberitakan satu kejadian yang sama dari macam macam berita online, untuk membaca beberapa dokumen yang membahas satu topik berita yang sama memerlukan waktu yang lumayan lama. Maka diperlukan suatu aplikasi yang dapat mengintegrasikan informasi dari beberapa dokumen berita. Peringkasan teks adalah proses untuk menyaring informasi yang paling penting dari sebuah atau beberapa sumber (dokumen) untuk menghasilkan versi yang lebih pendek yang berguna untuk *user* ataupun *task* tertentu.[7]

Karakteristik dari ringkasan adalah ringkasan yang bersifat informatif, dalam bentuk yang baik dan panjang dokumennya lebih pendek dari dokumen asli, artinya hasil ringkasannya masih mengandung informasi-informasi yang berkaitan dan penting dari teks asli dimana tidak terdapat pengulangan dari kalimat yang sama dan dapat dibaca oleh *user*. Dengan hasil dari aplikasi ini diharapkan pembaca terbantu dalam menyerap informasi penting yang ada dalam kumpulan dokumen berbeda dan tidak perlu lagi membaca dokumen satu per satu.

Terdapat dua pendekatan pada peringkasan teks, yaitu ekstraksi dan abstraksi. Pada teknik ekstraksi, sistem menyalin informasi yang dianggap paling penting dari teks asli menjadi ringkasan, contohnya pada klausa utama kalimat utama, atau paragraph utama. Sedangkan teknik abstraksi melibatkan parafrase dari teks asli. Pada umumnya, abstraksi dapat meringkas teks lebih kuat daripada ekstraksi, tetapi sistemnya lebih sulit dikembangkan karena mengaplikasikan teknologi *natural language generation* yang merupakan bahasan yang dikembangkan tersendiri. Metode *centroid-based* merupakan salah satu teknik untuk meringkas satu maupun beberapa dokumen yang menggunakan *centroid* dari klaster yang ada untuk menghasilkan ringkasan yang ekstraktif[9]. Pada aplikasi ini melibatkan proses *eliminasi stemming* dan *stopword* untuk mengetahui dan menganalisis pengaruh nya terhadap hasil ringkasan yang dihasilkan pada ringkasan ekstraktif. Untuk menghasilkan ringkasan diawali dengan beberapa dokumen yang relevan. Dokumen-dokumen tersebut akan dikelompokkan ke dalam klaster. *Centroid* dari klaster merupakan kumpulan dari *term* melalui proses perhitungan yang dianggap penting terhadap sebuah klaster dari dokumen-dokumen. Jenis *centroid* yang akan digunakan pada tugas akhir ini adalah *centroid* yang berupa *concept*. *Concept* dapat berupa kata atau frasa yang memiliki peranan semantik dalam sebuah kalimat[10]. Dengan *centroid* ini dapat diidentifikasi kalimat penting yang akan digunakan untuk membangun ringkasan. Untuk melakukan ekstraksi kalimat, dilakukan pembobotan terhadap kalimat-kalimat. Bobot dari kalimat adalah nilai yang menentukan seberapa penting kalimat tersebut dalam dokumen, semakin tinggi nilai bobot tersebut, maka kalimat tersebut memiliki prioritas yang lebih tinggi untuk menjadi bagian dari ringkasan. Adapun nilai-nilai bobot yang akan digunakan parameternya adalah *centroid value(C)*, *position value(P)*, *similarity*

with *tittle(F)* dan *Redundancy penalty(RS)*. Ke-empat parameter tersebut nantinya akan dihitung untuk menentukan ranking dari kalimat, kalimat dengan ranking tertinggi yang dihasilkan akan dijadikan kandidat untuk dijadikan sebagai ringkasan.

Studi kasus yang diambil adalah artikel berita, karena merupakan salah satu jenis dokumen teks yang banyak dibutuhkan orang dan dapat dengan mudah diperoleh di internet sebagai salah satu alternatif membaca berita dari koran. Melalui ringkasan, pembaca dapat mengetahui intisari atau informasi utama teks tanpa harus terlebih dahulu membaca keseluruhan teks sehingga dapat menghemat waktu pembacaan.

1.2 Perumusan Masalah

Dengan mengacu pada latar belakang di atas, maka permasalahan yang akan dibahas dan diteliti pada tugas akhir ini adalah:

1. Bagaimana menentukan kalimat-kalimat yang penting dengan menggunakan *centriod* dari klaster.
2. Bagaimana menghilangkan *redundancy data* dari hasil ringkasan yang didapatkan.
3. Bagaimana melakukan proses pengurutan (*ordering*) hasil dari ringkasan yang di peroleh.
4. Metode *stemming* apa yang akan digunakan untuk proses *eliminasi stemming*.
5. Bagaimana pengaruh pemotongan imbuhan kata (*stemming*) terhadap tingkat akurasi sistem yang dibangun.
6. Apakah penggunaan *stemming* dapat menghasilkan ringkasan yang lebih baik bila dibandingkan tanpa proses *stemming*.
7. Bagaimana pengaruh penghilangan kata-kata yang tidak penting (*stopword*) terhadap tingkat akurasi sistem yang dibangun.
8. Apakah penggunaan *stopword* dapat menghasilkan ringkasan yang lebih baik bila dibandingkan tanpa proses *stopword*.
9. Bagaimana mendapatkan kalimat dan kata-kata yang penting dalam melakukan proses pembobotan kalimat.
10. Bagaimana cara membuat bank dokumen untuk dokumen yang terlabeli dan untuk dokumen yang berulang
11. Melakukan evaluasi, apakah penerapan proses *stemming* dan *stopword* pada metode ini dapat menghasilkan ringkasan yang relevan dengan informasi pada teks asli.

Adapun batasan masalah pada tugas akhir ini yaitu:

1. Teks asli yang akan diringkas adalah dokumen berita berbahasa Indonesia.
2. Hasil peringkasan berupa ekstraksi terhadap kalimat dari teks asli.
3. Peringkasan yang dilakukan adalah peringkasan terhadap multi dokumen.
4. Inputan berupa suatu klaster dokumen dengan topik yang sama, dimana dokumennya sudah pernah digunakan pada tugas akhir tentang peringkasan teks di IT.Telkom.
5. Algoritma *stemming* yang di gunakan adalah algoritma *stemming* berbahasa Indonesia yaitu algoritma *stemmer* fadillah Z Tala.

6. Pada tugas akhir ini lebih menekankan pada analisis pengaruh nilai dan penggunaan *preprocessing stemming* dan *stopword* terhadap hasil ringkasan yang dihasilkan.
7. Untuk evaluasi, referensi ringkasan manusia juga berupa ringkasan ekstraksi.

1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah :

1. Membuat aplikasi peringkasan teks dengan metode *centroid-based* dengan mencoba melibatkan proses *eliminasi stemming* dan *stopword*.
2. Untuk mengetahui dan menganalisis seberapa besar pengaruh penggunaan *preprocessing stemming* dan penggunaan *stopword* terhadap ringkasan yang dihasilkan, dengan cara membandingkan dengan hasil ringkasan manusia dengan menggunakan *tool ROUGE*[2].
3. Mengimplementasikan metode tersebut menjadi sebuah sistem peringkasan otomatis untuk dokumen berita.
4. Hasil keluaran sistem yang diharapkan mengandung informasi utama yang relevan dari teks asli.

1.4 Metode Penyelesaian Masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah menggunakan metode studi pustaka atau studi literatur dan analisis dengan langkah kerja sebagai berikut:

1. Studi literatur
 - a. Pencarian referensi yang layak dan berhubungan dengan metode peringkasan teks, peringkasan dengan teknik *centroid based* dengan penggunaan *stemming* dan *stopword*.
 - b. Pendalaman materi, mempelajari dan memahami materi yang berhubungan dengan tugas akhir.
2. Analisis permasalahan
 - a. Mempelajari konsep *centroid based* untuk peringkasan teks yang akan digunakan dalam mengimplementasi perangkat lunak.
 - b. Menganalisis metode *centroid based* dengan melibatkan proses *preprocessing stemming* dan *stopword* untuk peringkasan teks dalam perancangan perangkat lunak, dengan menggunakan *tool ROUGE*.
3. Mengumpulkan *requirement* terhadap perangkat lunak yang akan dibangun.
4. Melakukan perancangan perangkat lunak yang akan dibangun.
5. Melakukan implementasi perancangan perangkat lunak.
6. Melakukan pengujian sistem dan menganalisis hasil keluaran sistem yang berupa ringkasan teks, sejauh mana ringkasan dapat menggambarkan makna utama teks dan mengujinya dengan *ROUGE evaluation toolkit*.
7. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.