BAB I
PENDAHULUAN

1.1. Latar Belakang

Dokumen teks digital memiliki informasi yang beraneka ragam dan dengan jumlah yang saat ini sudah semakin besar. Kadangkala juga dapat ditemukan beberapa atau banyak dokumen teks digital berbeda namun memiliki informasi yang hampir serupa. Untuk memudahkan dalam pengambilan informasi yang dperlukan, perlu adanya pengelempokan atau peringkasan dokumen-dokumen tersebut sehingga dapat diperoleh informasi dari dokumen tersebut tanpa harus membaca seluruhnya.

Multi Document Summarization adalah metode untuk menghasilkan suatu ringkasan tunggal dari beberapa dokumen yang memiliki kemiripan atau saling berhubungan. Metode ini sangat membantu dalam menyelesaikan masalah *information overload*, karena menghasilkan *document summary* / ringkasan dokumen untuk dibaca daripada harus membaca sekian banyak dokumen.

Pada penulisan tugas akhir ini, Peringkasan Multi Dokumen diimplementasikan menggunakan *Agglomerative Clustering* dan *Sentence Specific Feature Extraction*, yaitu metode yang melakukan tehnik peringkasan dengan melakukan *clustering* terhadap dokumen dan mengekstrak fitur-fitur kalimat dari dokumen tersebut.

Pada metode ini, *clustering* yang digunakan adalah metode *clustering hirarki agglomerative.* Metode ini dipilih karena dapat membentuk klaster-klaster dokumen dengan jumlah tertentu, sesuai dengan kebutuhan peringkasan dokumen / *document summarization*. Selain itu, diperkenalkan cara lain untuk menghitung bobot kata dengan menggunakan *Term Synonym Frequency - Inverse Sentence Frequency* (TSF – ISF). Bobot kata ini juga akan digunakan sebagai bagian dari ekstraksi *sentence specific feature* / fitur spesifik kalimat. Fitur spesifik kalimat ini kemudian digunakan untuk memperoleh skor kalimat / *sentence score.* Kalimat-kalimat yang memiliki skor tertinggi akan diekstrak menjadi ringkasan dokumen, sehingga ringkasan dokumen yang terbentuk akan menjadi ringkasan dokumen yang tidak kehilangan kalimat-kalimat yang menjadi pikiran utama dari dokumen yang diringkas.

1.2.Perumusan masalah

Dari latar belakang yang dipaparkan, permasalahan yang dirumuskan adalah :

1. Bagaimana mengimplementasikan metode *Agglomerative Clustering* dan Ekstraksi Fitur Spesifik Kalimat pada peringkasan multi dokumen untuk membangun membangun aplikasi peringkasan multi dokumen.
2. Bagaimana memberikan kondisi terminasi terbaik pada proses cluster yang dilakukan
3. Bagaimana mengukur kinerja aplikasi yang dibangun dalam melakukan peringkasan dokumen dari ringkasan dokumen yang dihasilkan

1.3.Batasan masalah

Adapun batasan masalah agar penelitian ini tidak meluas adalah sebagai berikut :

1. Dokumen yang akan diuji berupa file berekstensi *.txt
2. Sistem digunakan untuk dokumen bahasa Indonesia.
3. Isi teks bukan rumusan matematika dan bukan gambar.
4. Hasil peringkasan berupa ekstraksi terhadap kalimat dari teks dokumen asli.

1.4.Tujuan

Tujuan yang ingin dicapai adalah :

1. Mengimplementasikan metode *Agglomerative Clustering* dan Ekstraksi Fitur Spesifik Kalimat pada multi dokumen *summarization.*
2. Mencari kondisi terminasi terbaik pada proses *agglomerative clustering* yang dilakukan.
3. Mengukur kinerja aplikasi yang dibangun dalam melakukan peringkasan dokumen sesuai dengan nilai *compression rate* yang diberikan .

1.5. Metode Penyelesaian Masalah

Metodologi penyelesaian masalah yang akan digunakan adalah :

1. Studi literatur
   - Pencarian referensi dan sumber-sumber yang layak berkaitan dengan penyelesaian penelitian ini. Sumber referensinya meliputi materi yang berhubungan dengan *multi document summarization* melalui internet maupun buku, jurnal ilmiah atau orang yang berkompeten di bidang ini.
   - Pendalaman materi yang diperlukan berdasarkan referensi-referensi tersebut.
2. Pengumpulan data.

   Data terdiri dari sekumpulan dokumen teks yang dikumpulkan dari internet yang berekstensi *.txt.
3. Pembangunan Aplikasi
   - Analisis dan Perencanaan

     Membuat analisis serta perancangan aplikasi dalam mengimplementasikan *Agglomerative Clustering* dan Ekstraksi Fitur Spesifik Kalimat peringkasan multi dokumen.
   - Implementasi

     Dari hasil perencanaan yang telah dibuat maka akan diimplementasikan pada pemrograman komputer. Pemrograman dilakukan menggunakan bahasa pemrograman Java.
   - Pengujian

     Menguji hasil pemrograman dengan menggunakan skenario pengujian:

     Pengujian dilakukan dengan cara :

     Menghitung nilai F-Measure dari ringasan dokumen yang dihasilkan.

4. Analisis dan Pengambilan Kesimpulan dari Hasil Pengujian

   Membandingkan hasil dari aplikasi Peringkasan Multi Dokumen dengan melakukan uji coba terhadap aplikasi yang telah dibangun,. Kemudian hasilnya dievaluasi dengan menggunakan ROUGE *Evaluation toolkit* yang merupakan alat pengujian terhadap aplikasi peringkasan dokumen. ROUGE juga menggunakan tiga parameter, yaitu *precision, recall* dan *F-Measure*.