

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Peringkasan teks adalah proses merangkum informasi penting ke dalam bentuk yang lebih pendek tanpa menghilangkan informasi penting dari tulisan yang sebenarnya[1]. Peringkasan teks sangat penting untuk mempermudah pembaca dalam menyimpulkan informasi pada sebuah dokumen tanpa membaca keseluruhan isi dokumen, seperti peringkasan dokumen dengan metode *ViseKriterijumska Optimizacija I Kompromisno Resenje* (VIKOR) yang melakukan peringkasan dan pengurutan hasil ringkasan dokumen dengan memperhitungkan pengambilan keputusan berdasarkan multi kriteria seperti fitur posisi kalimat, kalimat positif, kalimat inti, keterkaitan kalimat dengan judul, Panjang relatif kalimat dan keterkaitan antar kalimat[2], yang kedua adalah metode Lexical Chain dan *Word Sense Disambiguation* yang melakukan peringkasan teks dengan proses pembentukan rantai leksikal berdasarkan kata-kata yang saling berkaitan, dimana setiap kata diperiksa ambiguitasnya untuk mendapatkan akurasi dalam pembentukan lexical chain yang sesuai dengan konteks kalimat[3] dan yang ketiga adalah metode *Fuzzy Logic* dan *Lexical Chains* yang melakukan peringkasan menggunakan proses ekstraksi fitur, pembentukan rantai leksikal dan logika fuzzy, dimana logika fuzzy berfungsi untuk memprediksi hasil ringkasan berdasarkan nilai-nilai parameter yang dikelompokkan berdasarkan nilai linguistik penting dan tidak penting[4].

Akan tetapi, sebagian besar hasil peringkasan teks terhadap beberapa dokumen dengan metode *extractive* sering tidak mempertimbangkan hubungan antar kalimat, sehingga kebanyakan informasi dari dokumen menjadi tidak tersampaikan, hal ini disebabkan karena pengurutan kalimat biasanya dilakukan berdasarkan urutan kemunculan kalimat pada dokumen. Berdasarkan Literature “*Sentence Ordering Using Cluster Correlation And Probability In Multi-*

Documents Summarization”, bahwa penggunaan *cluster correlation and probability* untuk pengurutan cluster pada dokumen berbahasa Inggris menghasilkan performansi hasil ringkasan yang lebih baik serta hasil ringkasan mudah dipahami dibandingkan pengurutan cluster menggunakan *cluster importance*[5].

Berdasarkan hal tersebut, pada penelitian ini akan mengimplementasikan dan mengukur performansi metode *cluster importance* dan *cluster correlation and probability* sebagai metode pengurutan cluster pada multi dokumen berbahasa Indonesia. Dengan adanya peringkasan teks otomatis yang memperhatikan hubungan antar kalimat, diharapkan dapat membantu pembaca untuk lebih mudah memahami makna ringkasan dari beberapa dokumen tanpa harus membaca keseluruhan isi dokumen.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang masalah pada penelitian ini terdapat permasalahan bagaimana menghasilkan peringkasan teks yang memperhatikan hubungan *rhetorical* antar kalimat, dengan menggunakan metode *cluster importance* dan *cluster correlation and probability* sebagai metode pengurutan cluster pada multi dokumen artikel berita berbahasa Indonesia.

1.3 Maksud dan Tujuan

Penelitian ini bermaksud membuat sistem peringkasan pada multi dokumen artikel berita berbahasa Indonesia. Sedangkan tujuan yang ingin dicapai dari penelitian ini adalah untuk mengukur performansi hasil ringkasan yang menggunakan metode *cluster importance* dan *cluster correlation and probability* sebagai metode pengurutan cluster.

1.4 Batasan Masalah

Pembatasan masalah dimaksudkan supaya dapat memberikan pemahaman yang terarah dan pembahasan tidak menyimpang dari pokok perumusan masalah yang ada, sehingga mencapai sasaran yang ditentukan sesuai dengan apa yang diharapkan. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dokumen yang diringkas adalah dokumen berita berbahasa Indonesia.
2. Topik dari data masukan adalah korupsi.
3. Masukan ke sistem berupa teks dengan format file .txt.
4. Dokumen masukan ke dalam sistem yang dibuat adalah lebih dari satu dokumen berita berbahasa Indonesia.
5. Jenis peringkasan yang digunakan adalah ekstraksi.
6. Hasil dari sistem peringkasan teks otomatis ini adalah teks ringkasan dan nilai performansi ringkasan.

1.5 Metode Penelitian

Metode penelitian yang akan digunakan dalam penelitian ini dapat dilihat pada gambar 1.1 berikut:



Gambar 1.1 Metode Penelitian

1. Identifikasi Masalah

Pada penelitian ini dilakukan identifikasi masalah dengan cara menganalisis penelitian sebelumnya yang sudah dilakukan. Salah satu hasil analisis adalah belum ada penelitian pengurutan hasil peringkasan

teks berdasarkan keterkaitan antar kalimat pada dokumen berbahasa Indonesia.

2. Perumusan Masalah

Tahapan ini merumuskan masalah yang didapat dari proses identifikasi masalah, dimana salah satu hasil analisis yang didapat adalah belum ada penelitian pengurutan hasil ringkasan teks menggunakan metode *cluster correlation and probability* pada dokumen berbahasa Indonesia.

3. Studi Literatur

Tahapan ini dilakukan untuk mengumpulkan berbagai informasi mengenai penelitian dengan cara mengumpulkan literatur, jurnal, buku, *ebook*, *paper* dan bacaan-bacaan yang berkaitan dengan topik penelitian yang akan dikerjakan.

4. Pengumpulan Data Masukan

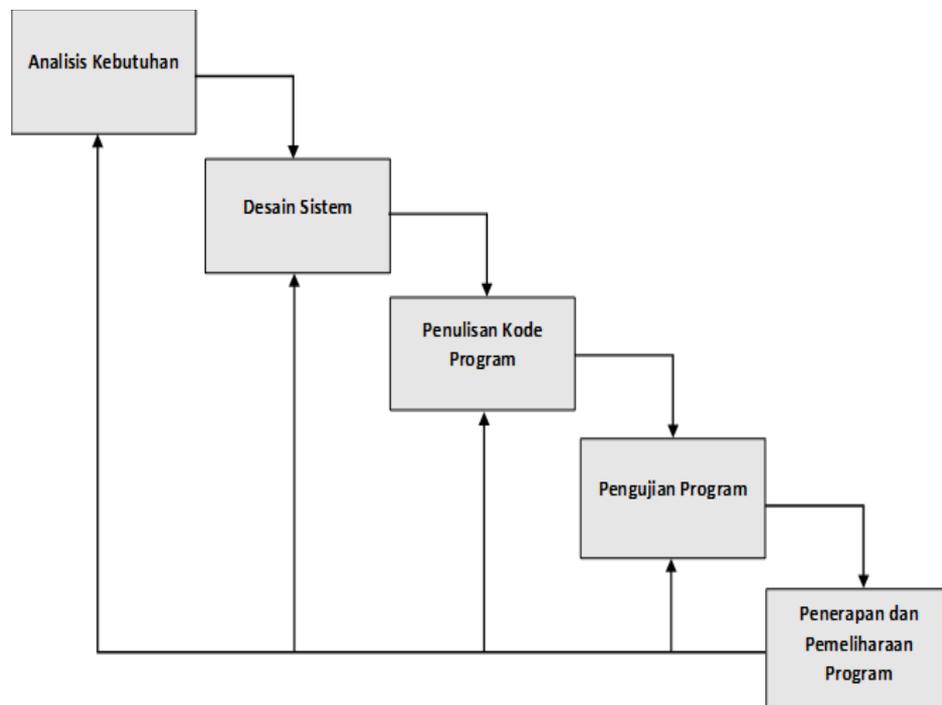
Pengumpulan data masukan dilakukan secara manual. Dimana data yang diambil merupakan berita dari portal pemuat berita seperti portal islam, tribunnews, tempo, okezone dan lain-lain. Data akan dipilih sesuai dengan topik, setelah itu disalin dan kemudian disimpan dalam format .txt. Data masukan dalam penelitian ini terdiri dari pengumpulan data masukan untuk proses pengujian.

5. Analisis Metode

Pada tahap ini dilakukan analisis peringkasan teks otomatis. Analisis yang dilakukan berupa analisis masalah, data masukan, *preprocessing*, clustering kalimat dengan metode *Similarity-based Histogram Clustering*, pengurutan cluster dengan metode *cluster importance* dan *cluster correlation serta probability*, ekstraksi kalimat dengan metode *distribution local sentence* dan pengukuran nilai performansi dengan metode *Recall Oriented Understudy for Gisting Evaluation*.

6. Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan untuk implementasi peringkasan teks multi dokumen berita berbahasa Indonesia ini menggunakan model *waterfall*, karena menggunakan pendekatan secara sistematis dan berurutan, sehingga cocok digunakan untuk sistem atau perangkat lunak yang bersifat generik, artinya sistem dapat diidentifikasi semua kebutuhannya dari awal dengan spesifikasi yang umum[6]. Model Pengembangan *Waterfall* dapat dilihat pada Gambar 1.2 berikut[7]:



Gambar 1.2 Model Pengembangan *Waterfall*

a. Analisis Kebutuhan

Tahap analisis kebutuhan ini bertujuan untuk memahami perangkat lunak yang diharapkan oleh pengguna dan batasan perangkat lunak tersebut. Pengumpulan informasi dilakukan untuk dianalisis supaya mendapatkan data yang dibutuhkan oleh pengguna.

b. Desain Sistem

Spesifikasi kebutuhan dari tahap sebelumnya akan dipelajari dalam tahap ini dan desain sistem disiapkan. Desain sistem membantu dalam penentuan perangkat keras, persyaratan sistem dan arsitektur sistem secara keseluruhan.

c. Penulisan Kode Program

Setelah desain sistem selesai, maka tahap selanjutnya adalah mengimplementasikannya ke dalam Bahasa pemrograman. Tahapan pertama adalah *preprocessing* seperti *segmentation*, *case folding*, *filtering*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya masuk ke dalam proses clustering kalimat dengan metode *Similarity-based Histogram Clustering*, dilanjutkan dengan proses pengurutan cluster dengan metode *cluster importance* dan *cluster correlation serta probability*, selanjutnya proses ekstraksi kalimat dengan metode *distribution local sentence* dan perhitungan performansi ringkasan.

d. Pengujian Program

Tahapan ini digunakan untuk pengujian program yang sudah dibangun. Apabila program yang dibuat berjalan sesuai rencana maka program peringkasan teks siap digunakan oleh pengguna sistem.

e. Penerapan dan Pemeliharaan Program

Pada tahapan ini sistem sudah siap digunakan, setelah itu dilakukan pengembangan sistem seperti penambahan fitur dan fungsi baru yang mungkin akan muncul untuk kebutuhan pengguna.

7. Pengujian Hasil Ringkasan Teks

Pada tahap ini dilakukan pengujian pada hasil ringkasan yang dihasilkan oleh sistem. Pengujian mengukur performansi hasil ringkasan yang menggunakan metode *cluster importance* dan *cluster correlation*

dan probability sebagai metode pengurutan cluster dengan menggunakan metode *Recall Oriented Understudy for Gisting Evaluation*.

8. Penarikan Kesimpulan

Pada tahap akhir ini dilakukan penarikan kesimpulan terhadap penggunaan metode pengurutan cluster yang telah di implementasikan untuk melakukan peringkasan teks otomatis.

1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mengenai penelitian yang akan dilakukan:

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang masalah, merumuskan masalah, menentukan maksud dan tujuan, batasan masalah, metode penelitian dan sistematika penulisan pada penelitian yang di angkat.

BAB 2 LANDASAN TEORI

Bab ini akan menjelaskan teori-teori pendukung yang berhubungan dengan penelitian ini yaitu mengenai peringkasan teks otomatis dengan metode *similarity-based histogram clustering*, *cluster correlation*, *cluster probability*, *cluster importance*, ekstraksi kalimat, clustering model dengan pemrograman berorientasi objek dan tools pendukung yang digunakan.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini akan membahas analisis dan perancangan dari sistem yang akan dibangun diantaranya analisis data masukan, analisis *preprocessing*, analisis metode *similarity-based histogram clustering* , analisis metode pengurutan cluster dengan analisis metode *cluster importance*, *cluster correlation and probability*, analisis data keluaran, analisis fungsional, analisis non fungsional dan perancangan antarmuka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini berisi mengenai implementasi dari analisis yang telah dilakukan, serta pengujian terhadap sistem peringkasan dan performansi hasil ringkasan multi dokumen berbahasa indonesia yang telah dibuat. Metode *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) digunakan untuk mengukur performansi hasil ringkasan.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi mengenai kesimpulan yang diperoleh dari hasil pengujian implementasi metode *cluster importance* dan *cluster correlation and probability* berdasarkan tujuan yang ingin dicapai serta saran dari masalah-masalah yang ditemukan selama proses penelitian untuk pengembangan sistem ke depan.