BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Kebanyakan mesin *information retrieval* (IR) modern menghasilkan daftar perangkingan dari dokumen, yang diukur dari relevansi dokumen dengan permintaan *user* (*user query*). Penaksiran pertama untuk mengukur hasil peringkasan yang relevan adalah dengan mengukur hubungan antar informasi yang ada dalam dokumen dengan *query* yang diberikan oleh user dan menambah kombinasi linier sebagai sebuah matrik. Kombinasi linier ini disebut "*marginal relevance*" [24].

Selain digunakan untuk *information retrieval*, *marginal relevance* bisa juga digunakan untuk peringkasan dokumen. Berdasarkan penelitian sebelumnya, metode *maximum marginal relevance* (MMR) sangat berguna untuk memberikan informasi pada dokumen yang bersifat multi redundansi dan dapat meningkatkan kecepatan pengukuran yang tepat dari kesamaan kalimat. Penelitian ini menghasilkan akurasi yang masih rendah (46%). Hal tersebut disebabkan karena *query* yang tidak terbatas, sehingga *user* bisa memasukkan kata apa saja pada *query* [3]. Pada penelitian lainnya, dengan data yang diambil dari surat kabar *online* berbahasa Indonesia (30 berita). Diperoleh hasil pengujian rata-rata *recall* 60%, *precision* 77%, dan *f-measure* 66% [25].

Idealnya panjang ringkasan itu tidak melebihi dari setengah teks asli [26]. Namun kenyataannya terkadang panjang ringkasan tidak sesuai dengan kriteria. Hal tersebut dikarenakan seluruh teks asli langsung diproses pembentukan ringkasan. Untuk mendapatkan hasil ringkasan yang sesuai dengan kriteria, dapat dilakukan pengurangan kandidat kalimat ringkasan. Pengurangan kandidat kalimat ringkasan dapat dilakukan dengan mengklasifikasikan kalimat ke dalam kelas relevan dan tidak relevan. Kalimat yang termasuk ke dalam kelas relevan akan dipilih dan diproses pada metode selanjutnya, yaitu dengan menghitung kesamaan kata antara kadidat kalimat ringkasan dengan *query*. Metode klasifikasi yang akan digunakan

untuk mengklasifikasikan kalimat pada penelitian ini adalah metode *K-Nearest Neighbors* (KNN). Pada penelitian sebelumnya, peringkasan dokumen dengan metode *K-Nearest Neighbors*, diperoleh hasil dengan nilai *recall* 63%, *precision* 100%, dan *f-measure* 77% [4].

Berdasarkan uraian di atas, pada penelitian ini akan dilakukan *Text Summarization* menggunakan metode *K-Nearest Neighbors* dan *Maximum Marginal Relevance*, untuk meringkas sebuah artikel berbahasa Indonesia.

1.2. Identifikasi Masalah

Berdasarkan uraian latar belakang masalah yang telah dijelaskan sebelumnya, maka identifikasi masalah adalah sebagai berikut.

1. Masih rendahnya akurasi peringkasan menggunakan MMR
2. Masih panjangnya ringkasan yang dihasilkan

1.3. Maksud dan Tujuan

1.3.1. Maksud

Penelitian ini bermaksud membuat sistem peringkas teks otomatis.

1.3.2. Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah untuk membandingkan nilai performansi (*recall, precision* dan *f-measure)* dan akurasi peringkasan menggunakan KNN dan MMR dengan peringkasan yang hanya menggunakan MMR.

1.4. Batasan Masalah

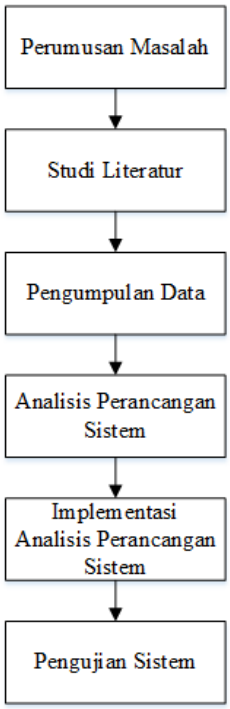
Batasan masalah merupakan pembatasan ruang lingkup pekerjaan yang akan dilakukan terhadap permasalahan yang ada. Batasan dalam penelitian adalah sebagai berikut:

1. Dokumen yang akan diringkas adalah dokumen berbahasa Indonesia
2. Dokumen yang digunakan bersumber dari beberapa media internet (viva.co.id, detik.com, kompas.com, dan merdeka.com)
3. Input teks berupa satu dokumen

4. Dokumen yang digunakan mempunyai judul yang sesuai dengan isi dokumen

1.5. Metodologi Penelitian

Penelitian ini menggunakan Metode Deskriptif [27] yang bertujuan mendapatkan gambaran yang jelas mengenai hal-hal yang diperlukan. Dalam penelitian *text summarization* ini, data yang diteliti dalam keadaan sebagaimana adanya. Sehingga hasil *text summarization* tergantung dari data itu sendiri.



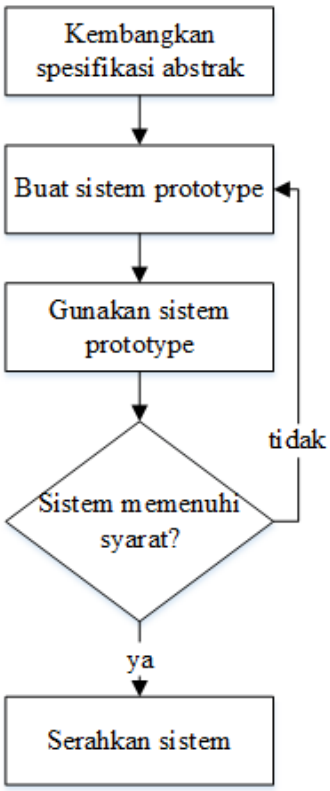
Gambar 1. 1 Metodologi Penelitian

1.5.1. Metode Pengumpulan Data

Pada penelitian ini metode yang digunakan adalah studi literatur. Studi literatur yang diperoleh dari sumber-sumber tertulis, baik yang tercetak maupun elektronik, seperti buku, *e-book,* jurnal, *paper,* dan sumber-sumber yang berhubungan dengan permasalahan yang diteliti.

1.5.2. Metode Pembangunan Sistem

Metode pembangunan perangkat lunak yang digunakan adalah *prototype* [5]. Karena pada pembangunan sistem *text summarization* didasarkan pada ide pengembangan awal, diuji keberhasilannya, dan menyempurnakannya dengan kembali ke pembangunan. Tahapan-tahapan dari metode *prototype* dapat dilihat pada Gambar 1.2.



Gambar 1. 2 Metode *Prototype* [5]

1. Kembangkan spesifikasi abstrak

   Merupakan tahap awal untuk menganalisis masalah yang ada, yaitu memahami masalah yang timbul dan mencari solusi untuk memecahkan masalah dalam menghasilkan sebuah *text summarization* dari artikel berbahasa Indonesia.

2. Buat sistem *prototype*

   Merupakan tahap pembuatan sistem *prototype* yang sudah didapatkan dari hasil pengumpulan kebutuhan dan perbaikan. Pada tahap ini pengembang

mencoba mengimplementasikan metode MMR pada *text summarization* ke dalam logika-logika program.

3. Gunakan sistem *prototype*

   Program akan diuji, dimana uji coba dilakukan untuk mengetahui kekurangan pada program. Jika masih ada kekurangan, maka p*rototype* direvisi dengan tahapan-tahapan yang sebelumnya telah dilakukan. Proses ini akan terus berulang sampai *prototype* yang dihasilkan mendekati harapan yang diinginkan.

4. Serahkan sistem

   Jika hasil sistem *prototype* telah mencapai atau mendekati harapan yang diinginkan. Maka sistem *text summarization* menggunakan KNN dan MMR bisa digunakan dengan baik.

1.6. Sistematika Penulisan

Sistematika penulisan laporan akhir penelitian *text summarization* disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang urutan pemahaman dalam menyajikan laporan penelitian *text summarization.*

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai *text summarization*, *preprocessing*, metode *Term Frequency – Inversed Document Frequency* (TF-IDF), *cosine similarity*, K-*Nearest Neighbors* (KNN), *Maximum Marginal Relevance* (MMR), bahasa pemrograman, *software* pendukung, dan model pernagkat lunak.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini menguraikan penjelasan analisis dan perancangan sistem, perhitungan pada metode *Term Frequency – Inversed Document Frequency* (TF-IDF), *cosine similarity*, K-*Nearest Neighbors* (KNN), *Maximum Marginal Relevance* (MMR) untuk diaplikasikan pada *Text Summarization.*

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang implementasi dari analisis dan perancangan yang telah dilakukan pada bab sebelumnya. Implementasi yang dilakukan terdiri dari implementasi perangkat keras, implementasi perangkat lunak, implementasi basis data dan implementasi antarmuka. Selain itu dilakukan pengujian *Text Summarization* menggunakan metode K-*Nearest Neighbors* (KNN) dan *Maximum Marginal Relevance* (MMR) pada Artikel Berbahasa Indonesia, dengan skenario pengujian dan hasil pengujian.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari penelitian *Text Summarization* menggunakan metode K-*Nearest Neighbors* (KNN) dan *Maximum Marginal Relevance* (MMR pada artikel berbahasa Indonesia, serta saran–saran untuk pengembangan penelitian selanjutnya.