# BAB 1
# PENDAHULUAN

## 1.1 Latar Belakang Masalah

Peringkasan teks otomatis (*automatic text summarization*) adalah pembuatan versi yang lebih singkat dari sebuah teks dengan memanfaatkan aplikasi yang dijalankan pada komputer. Hasil peringkasan ini mengandung poin-poin penting dari teks asli. Ringkasan didefinisikan sebagai sebuah teks yang dihasilkan dari satu atau lebih teks, mengandung informasi dari teks asli dan panjangnya tidak lebih dari setengah teks asli [1].

Terdapat dua pendekatan pada peringkasan teks, yaitu ekstraksi (*shallower approaches*) dan abstraksi (*deeper approaches*). Pada teknik ekstraksi, sistem mengambil kata kata penting yang akan disusun menjadi sebuah ringkasan. 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 [1].

Berdasarkan jumlah sumbernya, ringkasan teks dapat dihasilkan dari satu sumber (*single-document*) atau dari banyak sumber (*multi-document*). Suatu ringkasan dapat bersifat general, yaitu ringkasan yang berupaya mengambil sebanyak mungkin informasi umum yang mampu menggambarkan keseluruhan isi teks [1].

Pada penelitian peringkasan multi dokumen yang dilakukan oleh Tendi Arifin didapatkan bahwa hasil dari akurasi sebesar 67.35%, *recall* sebesar 47.83%, *precission* sebesar 35.48, dan *f-measure* sebesar 40.74% [2]. Setelah diperoleh f-measure, ternyata metode RVM masih kurang baik performansinya sehingga diperlukan metode yang lain untuk meningkatkan akurasi pada peringkasan multi dokumen.

Penelitian yang dilakukan oleh Yuzar Marsyah menyatakan bahwa hasil ringkasan algoritma textrank mengandung kalimat yang lebih relevan dengan hasil 95.56% [3]. Algoritma TextRank juga lebih cepat dalam mengeluarkan hasil ringkasan dengan 0.0294 detik dibandingkan algoritma LexRank dengan 0.4321

detik [3]. Pada penelitian yang dilakukan oleh Ailin Li, Tao Jiang, Qingshuai Wang dan Hongzhi Yu tentang peringkasan teks dokumen tunggal menggunakan textrank dan lexrank didapatkan hasil akurasi sebesar 80% [4]. Dengan begitu akan diuji akurasi algoritma *Textrank* pada multi dokumen

Berdasarkan penjelasan diatas, tugas akhir ini mengimplementasikan Algoritma TextRank dalam peringkasan teks otomatis pada multi dokumen menggunakan Bahasa Indonesia

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, dapat diidentifikasi masalah yaitu belum diketahuinya akurasi dan performansi dari algoritma textrank untuk peringkasan teks multi dokumen.

## 1.3 Maksud dan Tujuan

### 1.3.1 Maksud

Maksud dari penelitian ini adalah untuk mengimplementasikan algoritma textrank dalam pembuatan peringkasan teks otomatis pada multi dokumen.

### 1.3.2 Tujuan

Adapun tujuan dari penelitian ini adalah mengukur akurasi dan performansi peringkasan teks otomatis pada multi dokumen dengan menggunakan algoritma textrank.

## 1.4 Batasan Masalah

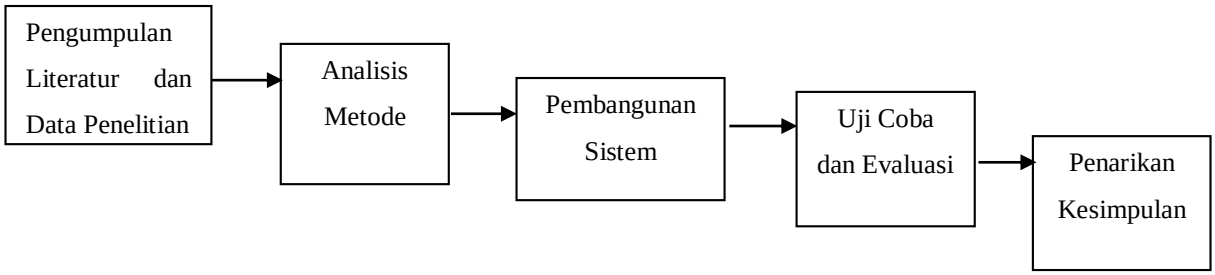
Batasan masalah merupakan pembatasan ruang lingkup pekerjaan yang akan dilakukan terhadap permasalahan yang ada. Batasan dalam penelitian adalah sebagai berikut:

1. Peringkasan yang dilakukan adalah peringkasan terhadap multi dokumen berita berbahasa Indonesia

2. Metode peringkasan yang digunakan adalah metode Ekstrasi
3. Dalam inputan, berisi dokumen-dokumen yang dikelompokkan dalam topik yang sama.
4. Pemodelan analisis yang digunakan adalah pemodelan analisis berorientasi objek.
5. Dokumen masukan berformat .txt

## 1.5 Metode Penelitian

Metode penelitian yang akan digunakan dalam penelitian ini dapat dilihat pada Gambar 1.1




**Gambar 1. 1 Metode Penelitian**

Adapun detail dari metode penelitian pada Gambar 1.1 adalah sebagai berikut.

a. Pengumpulan Literatur dan Data Penelitian

Pengumpulan data dilakukan dengan cara mengumpulkan berbagai literatur dan referensi yang berkaitan dengan penelitian yang dilakukan, mulai dari buku, jurnal, dan artikel. Selain itu dilakukan juga pengumpulan dokumen-dokumen berbahasa Indonesia yang akan dijadikan bahan pengujian dari penelitian yang akan dilakukan.

Setelah itu melakukan tinjauan literatur (*literature review*). Tinjauan literatur berisi uraian tentang teori, temuan dan bahan penelitian lain yang diperoleh dari bahan acuan untuk dijadikan landasan kegiatan penelitian.

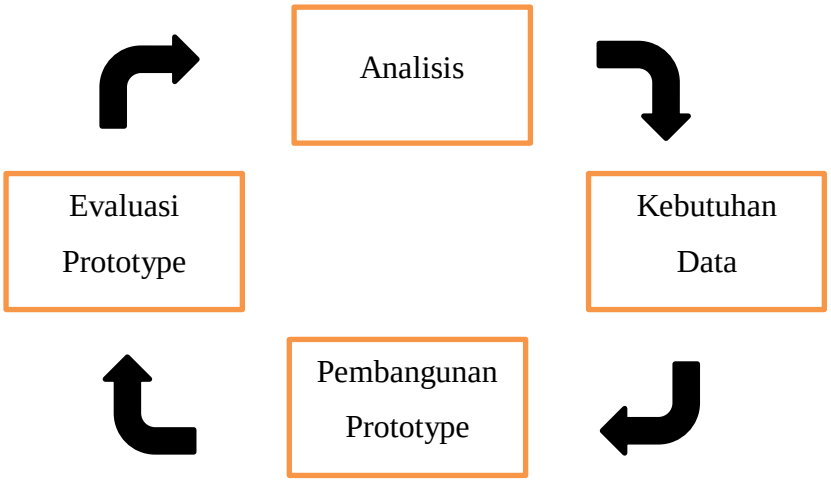
Tinjauan literatur dilakukan dengan cara membaca beberapa referensi-referensi pendukung penelitian yang telah dikumpulkan sebelumnya, lalu meringkas referensi-referensi tersebut sehingga membentuk pemahaman yang lebih mendalam terhadap hal-hal yang akan diteliti.

b. Analisis Metode

Dalam penelitian ini analisis dibagi menjadi beberapa bagian, yaitu analisis masalah yang diangkat, analisis data masukan, analisis tahap preprocessing, analisis peringkasan multi dokumen dengan melakukan penghitungan manual dari metode yang digunakan, serta analisis kebutuhan fungsional dan non fungsional dari perangkat lunak yang akan dibangun. Setelah analisis dilakukan, dibuat suatu perancangan model perangkat lunak yang sebelumnya telah dianalisis kebutuhan fungsionalnya, dimana perangkat lunak yang dibangun untuk mensimulasikan metode yang diterapkan.

c. Pembangunan Sistem

Setelah melakukan tahapan analisis algoritma textrank yang akan digunakan untuk peringkasan teks otomatis, langkah selanjutnya adalah implementasi metode tersebut dengan cara membangun perangkat lunak menggunakan bahasa pemrograman Java. Implementasi pembangunan sistem peringkasan teks multi dokumen menggunakan metode prototype. Pembangunan sistem peringkasan teks akan selalu dikembangkan.




**Gambar 1. 2 Metode Prototype**

1. Analisis

Pada tahap ini dilakukan analisis dari studi literatur penelitian sebelumnya yang telah dikerjakan, analisis data yang akan menjadi masukan hingga analisis metode TextRank

2. Kebutuhan Data

Pada tahap ini dilakukan pengumpulan data yang dibutuhkan sistem, dimana data dari sistem ini didapat dari situs liputan6.com

3. Pembangunan *prototype*

Pada tahap ini dilakukan implementasi dari proses analisis metode yang telah dilakukan sebelumnya dan perancangannya kedalam sistem. Pada tahap ini mengimplementasikan metode TextRank.

4. Evaluasi Prototype

Program akan diuji untuk mendapatkan kekurangan dari program. Jika masih ada kekurangan, maka akan diperbaiki kembali melalui tahapan-tahapan yang ada pada metode *prototype*. Proses ini akan diulang sampai prototype yang dihasilkan mendekati ekspetasi yang diinginkan.

d. Uji Coba dan Evaluasi

Melakukan uji coba dengan menggunakan perangkat lunak yang telah dibuat. Uji coba yang dilakukan adalah dengan melakukan proses pengujian dengan menggunakan dokumen-dokumen uji yang telah disiapkan, melihat akurasi dari setiap percobaan, serta membuat kesimpulan terhadap hasil percobaan tersebut apakah algoritma textrank memiliki akurasi yang baik atau tidak.

## 1.6 Sistematika Penulisan

Sistematika penulisan tugas akhir penelitian ini 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, rumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian serta sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang urutan pemahaman dalam menyajikan laporan ini.

**BAB 2 LANDASAN TEORI**

Bab ini membahas mengenai landasan teori yang digunakan untuk menganalisis masalah dan teori yang dipakai dalam data penelitian yaitu teori mengenai *peringkasan teks otomatis pada multi dokumen, pre-processing,* algoritma textrank, bahasa pemograman yang digunakan.

**BAB 3 ANALISIS DAN PERANCANGAN**

Bab ini menguraikan penjelasan analisis dan perancangan sistem, perhitungan pada algoritma textrank untuk diaplikasikan pada *peringkasan teks otomatis pada multi dokumen.*

**BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Bab ini berisi tentang pembangunan perangkat lunak, pengujian sistem dan pengujian akurasi algoritma textrank

**BAB 5 KESIMPULAN DAN SARAN**

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil *peringkasan teks pada multi dokumen m*enggunakan algoritma textrank Pada multi dokumen skripsi serta saran–saran untuk pengembangan media pembelajaran kedepannya.