

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Text mining atau penambangan teks adalah sebuah proses mengekstraksi pola berupa informasi dan pengetahuan yang berguna dari sumber data teks seperti Dokumen *Word*, *PDF*, Kutipan, dan lain-lain. Menurut Ronen Feldman dan James Sanger, *Text Mining* dapat didefinisikan secara luas sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah kategorisasi [1]. Dalam *text mining* mengembangkan banyak teknik pada bidang lain seperti, *information retrieval*, *machine learning*, *text summarization*, *natural language processing*, dan lain-lain.

Information retrieval atau temu-kembali informasi merupakan salah satu bagian dari *text mining*. Menurut Gerald J. Kowalski, temu-kembali informasi adalah suatu sistem yang mampu melakukan penyimpanan pencarian dan pemeliharaan informasi (teks, gambar, audio, video) [2]. Penelitian tentang temu-kembali informasi sudah pernah dilakukan sebelumnya oleh Tamara G Kolda, dalam penelitian tersebut menggunakan *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition*. Dari penelitian tersebut didapatkan kesimpulan metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dapat diterapkan dan menghasilkan kesimpulan yang baik pada temu-kembali informasi [3]. Hasil dari penelitian sendiri tidak dijelaskan nilai akurasi yang didapatkan.

Selain temu-kembali informasi, *text summarization* atau peringkasan teks merupakan teknik lain dari *text mining*. Peringkasan teks adalah memendekan, mengikhtisarkan, mengambil inti dari sebuah teks/dokumen. Dalam penelitian ini akan menggunakan metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* pada peringkasan teks untuk menguji metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dapat diimplementasikan pada

peringkasan teks yang merupakan bagian dari *text mining* seperti temu-kembali informasi.

Studi kasus dalam penelitian ini adalah latar belakang proposal skripsi. Proposal skripsi adalah proposal yang diajukan oleh mahasiswa kepada dosen pada saat memulai melakukan skripsi. Tetapi dalam pembacaan proposal membutuhkan waktu yang lama, sehingga dibutuhkan bantuan peringkasan teks otomatis agar dosen dapat mengetahui inti dari latar belakang proposal skripsi tersebut.

Berdasarkan uraian diatas, maka dalam penelitian ini membangun peringkasan teks otomatis menggunakan metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* pada latar belakang proposal skripsi. Hasil penelitian ini diharapkan akan menghasilkan hasil yang baik seperti pada penelitian temu-kembali informasi yang pernah dilakukan sebelumnya.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang masalah, dapat dirumuskan masalah dalam penelitian ini adalah apakah metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dapat digunakan dalam peringkasan teks otomatis.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah menerapkan metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* pada peringkasan teks otomatis latar belakang proposal skripsi. Sedangkan tujuan dari penelitian ini adalah mengukur akurasi metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dalam peringkasan latar belakang proposal skripsi.

1.4 Batasan Masalah

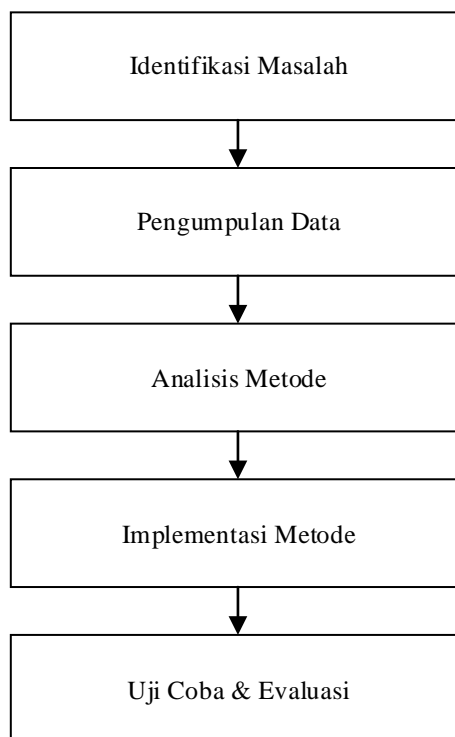
Adapun beberapa batasan-batasan masalah dalam penelitian ini yaitu sebagai berikut:

1. Dokumen yang diringkas adalah dokumen teks berbahasa indonesia.
2. Dokumen yang diringkas adalah dokumen tunggal (*single document*).
3. Teks masukan berupa latar belakang proposal skripsi yang sudah divalidasi oleh ahli bahasa.

4. Jenis peringkasan yang digunakan adalah *ekstraktif*.
5. Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan dalam pembobotan *term*.
6. Metode *Cross Method LSA* digunakan dalam menentukan kalimat ringkasan.
7. Hasil ringkasan sistem hanya diambil paling besar 50% dari jumlah kalimat pada latar belakang proposal skripsi.

1.5 Metode Penelitian

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



Gambar 1.1 Metode Penelitian

1.5.1 Identifikasi Masalah

Identifikasi masalah dengan cara menganalisis penelitian sebelumnya yang sudah dilakukan dimana salah satu hasil dari analisis adalah penelitian menggunakan metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dapat diterapkan pada temu-kembali informasi. Tetapi belum ada penelitian tentang peringkasan teks menggunakan metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition*. Sehingga belum diketahui apakah metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dapat diimplementasikan sama baiknya pada peringkasan teks seperti pada temu-kembali informasi.

1.5.2 Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengumpulkan berbagai literatur dan referensi yang berkaitan dengan penelitian yang dilakukan, mulai dari buku, jurnal, artikel, dan bacaan-bacaan yang ada kaitannya dengan peringkasan teks. Selain itu dilakukan juga pengumpulan latar belakang proposal skripsi yang akan dijadikan sebagai pengujian peringkasan.

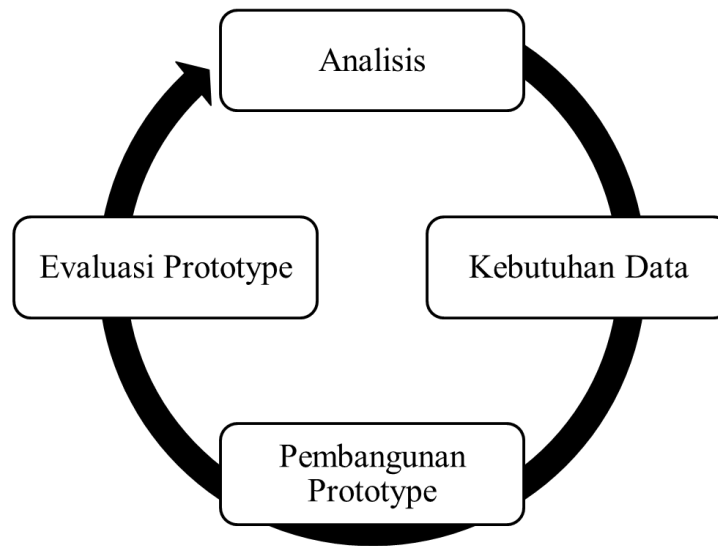
1.5.3 Analisis Metode

Analisis metode dibagi menjadi beberapa bagian, yaitu analisis data masukan, analisis tahap *Preprocessing*, analisis metode *Term Frequency-Inverse Document Frequency*, analisis metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition*, analisis metode *Cross Method LSA*, serta analisis kebutuhan *fungsiional* dan *non-fungsiional* dari perangkat lunak yang akan dibangun.

1.5.4 Implementasi Metode

Implementasi metode dengan cara membangun sistem peringkasan teks otomatis menggunakan bahasa pemrograman *Python*. Implementasi dilakukan berdasarkan analisis dan perancangan yang telah dilakukan sebelumnya dari mulai tahap *Preprocessing*, pembobotan *Term Frequency-Inverse Document Frequency*, implementasi *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition*, dan *Cross Method LSA*.

Implementasi pembangunan perangkat lunak peringkasan latar belakang proposal skripsi menggunakan metode *Prototype*. Karena pada pembangunan sistem peringkasan teks terus menerus diperbaiki [4].



Gambar 1.2 Metode Prototype

1. Analisis

Tahap ini dilakukan analisis dari studi literatur pada penelitian yang pernah dilakukan sebelumnya, analisis data yang akan menjadi masukan, analisis *Preprocessing*, analisis pembobotan *Term Frequency-Inverse Document Frequency*, analisis metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* dalam mendekomposisi, dan analisis metode *Cross Method LSA* dalam pemilihan kalimat yang akan dijadikan ringkasan.

2. Kebutuhan Data

Tahap ini dilakukan pengumpulan data dan melakukan validasi dari data yang dikumpulkan. Data diambil dari latar belakang proposal skripsi. Data divalidasi oleh pakar untuk dapat dijadikan data masukan.

3. Pembangunan *Prototype*

Tahap implementasi dari proses analisis metode dan perancangan sistem yang sudah didapatkan. Pada tahap ini mengimplementasikan tahapan *Preprocessing*, metode *Term Frequency-Inverse Document Frequency* dalam pembobotan kata, metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* untuk peringkasan teks. Dan metode *Cross Method LSA* dalam pemilihan kalimat ringkasan.

4. Evaluasi *Prototype*

Tahap pengujian sistem, dimana pengujian dilakukan untuk mengetahui kekurangan pada sistem. Jika masih ditemukan kekurangan dalam tahapan pembangunan, maka akan diperbaiki kembali melalui tahapan-tahapan yang ada pada metode *Prototype*. Proses ini akan diulang sampai *Prototype* yang dihasilkan dapat melakukan seluruh tahapan yang ada pada pembangunan *Prototype*.

1.5.5 Uji Coba dan Evaluasi

Melakukan uji coba dengan menggunakan sistem yang telah dibangun. Uji coba dilakukan adalah dengan melakukan proses pengujian dengan menggunakan dokumen latar belakang proposal skripsi yang telah disiapkan, melihat akurasi dari setiap percobaan, serta membuat kesimpulan terhadap hasil percobaan tersebut apakah metode *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition* memiliki hasil yang baik atau tidak.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Adapun sistematika penulisan yang dilakukan dalam penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini membahas secara umum mengenai latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, metode pembangunan perangkat lunak, dan sistematika penulisan yang dimaksud agar

dapat memberikan gambaran tentang urutan penulisan skripsi peringkasan teks otomatis

BAB 2 LANDASAN TEORI

Bab ini membahas secara umum teori-teori dalam skripsi peringkasan teks otomatis adapun teori-teori yang digunakan yaitu peringkasan teks otomatis, *Preprocessing*, *Term Frequency Inverse Document Frequency*, *Latent Semantic Analysis*, *Semi Discrete Decomposition*, *Cross Method LSA*, evaluasi ringkasan, *Unified Modeling Language*, pengujian *Black Box*.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini menguraikan penjelasan analisis data masukan, analisis *Preprocessing*, analisis pembobotan *Term Frequency Inverse Document Frequency*, Analisis *Latent Semantic Analysis* berbasis *Semi Discrete Decomposition*, analisis *Cross Method LSA*, analisis kebutuhan *nonfungsional* dan *fungsional*, dan analisis perancangan sistem.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini menjelaskan tentang implementasi dan pengujian dari sistem yang dibangun. Implementasi dan pengujian yang dilakukan berupa implementasi perangkat keras, implementasi perangkat lunak, implementasi antarmuka, dan implementasi prototype.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dari hasil penelitian yang telah dilakukan dan saran yang diharapkan dapat dikembangkan untuk penelitian selanjutnya.

