BAB 1
PENDAHULUAN

1.1. Latar Belakang

Dengan semakin pesatnya perkembangan teknologi maka kebutuhan untuk mendapatkan informasi semakin besar. Informasi sekarang bisa di dapatkan secara cepat dan mudah serta bisa di dapatkan kapan dan dimana saja melalui berbagai macam portal berita yang tersedia untuk memenuhi kebutuhan masyarakat akan informasi yang *up to date* dan *real time,* tetapi dari sekian banyak artikel berita yang disediakan oleh portal – portal berita tersebut tidak mempunyai ringkasan dan abstrak. Untuk mendapatkan informasi inti dari sebuah artikel berita, pembaca harus membaca keseluruhan artikel yang cukup memakan waktu, sehingga dibutuhkan sebuah alat bantu yang dapat menghasilkan ringkasan dari dokumen sumber secara otomatis dan akurat.

Peringkas teks otomatis merupakan sebuah alternatif yang dapat digunakan untuk menghasilkan ringkasan secara otomatis menggunakan bantuan komputer. Pada penerapannya terdapat banyak metode yang digunakan untuk menghasilkan ringkasan pada peringkas teks otomatis. Salah satu metode yang digunakan untuk menghasilkan ringkasan adalah metode *Latent Semantic Analysis* (LSA). Metode ini terinspirasi oleh metode *Latent Semantic Indexing* (LSI) yang menerapkan *Singular Value Decomposition* (SVD) sebagai cara untuk memilih kalimat yang akan dijadikan sebagai kalimat ringkasan [1].

Setelah metode ini dikenalkan pada tahun 2001, banyak terjadi pengembangan terhadap metode ini untuk menghasilkan ringkasan yang lebih akurat yaitu : penelitian yang dilakukan oleh Steinberg dan Jezek pada 2004 [2] yang perbaikan terhadap tingkat akurasi hasil ringkasan pada metode *latent semantic analysis*, kemudian penelitian yang dilakukan oleh Murray,Renals dan Carletta pada tahun 2005 [3] yang melakukan proses modifikasi pada tahapan reduksi matriks serta modifikasi pada tahap pengambilan ringkasan yang menghasilkan ringkasan yang lebih akurat dibanding dengan metode sebelumnya, serta yang terbaru adalah penelitian yang dilakukan oleh Makblu,Ilyas dan Ferda

pada tahun 2010 [4] yang menghasilkan dua buah metode latent semantic yang baru yaitu *Topic Method* dan *Cross Method Latent Semantic Analysis*.

Sedangkan penerapan metode LSA untuk meringkas teks berbahasa Indonesia juga banyak bermunculan, seperti penggabungan metode LSA dengan metode *clustering* dokumen untuk menghasilkan ringkasan pada *multi document* berbahasa Indonesia [5] yang berhasil meningkatkan tingkat akurasi yang dihasilkan dari multi document dengan cara menggabungkan algoritma *Latent Semantic Analysis* dengan algoritma *clustering* dan pada penelitian lainnya digunakan metode *Latent Semantic Analysis* dan metode *Relevance Measure* untuk meringkas dokumen tunggal berbahasa Indonesia [6] yang membuktikan bahwa metode *Latent Semantic Analysis* mempunyai tingkat akurasi ringkasan yang lebih tinggi dibanding dengan metode *Relevance Measure*.

Dari penelitian yang menggunakan metode *latent semantic analysis* untuk menghasilkan ringkasan pada teks berbahasa Indonesia diketahui bahwa metode *latent semantic analysis* yang digunakan adalah metode *latent semantic analysis* awal yang belum dikembangkan. Penggunaan *metode latent semantic analysis* yang lama tentu akan menghasilkan ringkasan dengan tingkat akurasi yang lebih rendah jika dibandingkan dengan akurasi ringkasan yang dihasilkan dari metode *latent semantic analysis* yang telah dikembangkan.

Sebuah Peringkas teks otomatis diharapkan dapat membantu pembaca untuk mendapatkan informasi inti dari berita yang dibaca dengan lebih cepat. Penerapan *Cross Method Latent Semantic Analysis* pada peringkas teks otomatis dapat menghasilkan sebuah ringkasan yang lebih akurat dibanding dengan ringkasan yang dihasilkan metode *latent semantic analysis* biasa [4]. Oleh karena itu, penulis tertarik untuk mengambil pokok bahasan penelitian dengan judul "Implementasi *Cross Method Latent Semantic Analysis* Pada peringkas Teks Otomatis Untuk Artikel Berita Berbahasa Indonesia".

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dari penelitian ini adalah bagaimana cara untuk mengimplementasikan *Cross*

Method Latent Semantic Analysis pada peringkas teks otomatis untuk artikel berita berbahasa Indonesia untuk menghasilkan ringkasan yang cepat dan akurat.

1.3. Maksud dan Tujuan Penelitian

Berdasarkan permasalahan yang telah di uraikan, maka maksud dari penelitian ini adalah untuk melakukan implementasi *Cross Method Latent Semantic Analysis* pada peringkas teks otomatis.

Sedangkan tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi dari ringkasan yang dihasilkan dengan penerapan *Cross Method Latent Semantic Analysis* pada peringkas teks otomatis terhadap artikel berita berbahasa Indonesia.

1.4.Batasan Masalah

Untuk menghindari pembahasan yang meluas, maka penulis membatasi permasalahan sebagai berikut :

a. Dokumen yang di ringkas adalah artikel berita politik berbahasa Indonesia.
b. Dokumen yang dimasukan kedalam peringkas teks otomatis diambil menggunakan *rich site summary* (RSS) yang dimiliki portal berita viva.co.id yang merupakan salah satu portal berita dengan tingkat akses pengunjung tinggi.
c. Jumlah Berita yang ditampilkan oleh peringkas teks otomatis dari portal berita viva.co.id yang dapat dipilih oleh pengguna untuk di ringkas adalah 15 buah berita.
d. Dokumen yang di masukan ke dalam peringkas teks otomatis untuk di ringkas merupakan dokumen tunggal.
e. Hasil yang dihasilkan oleh peringkas teks otomatis adalah ringkasan.
f. Bahasa program yang akan digunakan adalah PHP hypertext preprocessor (PHP) dan MySQL sebagai pengolahan *database.*
g. Data *stop words* bahasa Indonesia yang digunakan diperoleh dari Kevin Bouge[1].
h. Data kata dasar bahasa Indonesia yang digunakan diperoleh dari kateglo[2]

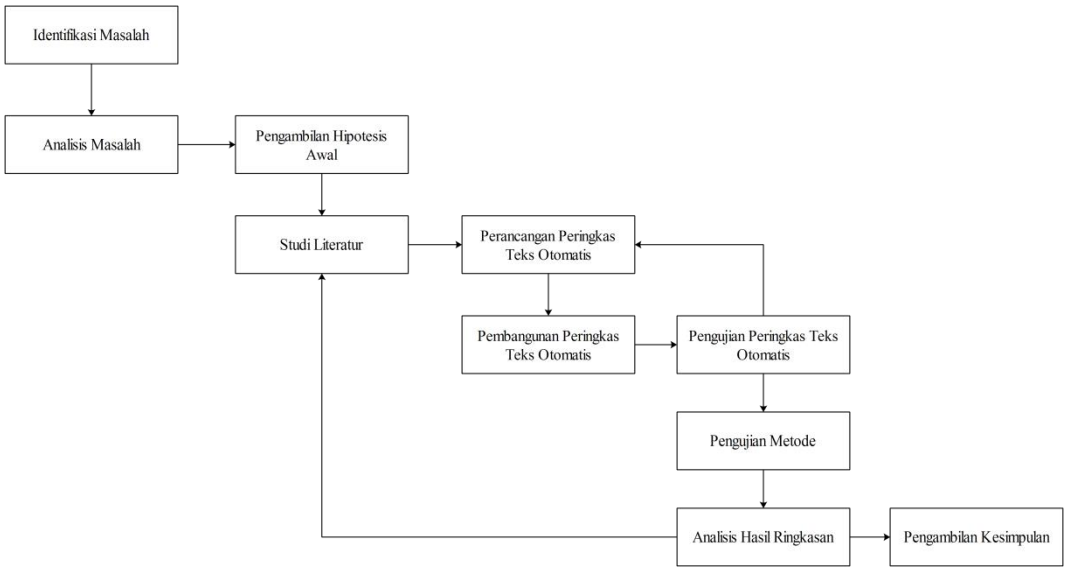
[1] https://sites.google.com/site/kevinbouge/stopwords-lists

[2] https://github.com/ivanlanin/kateglo

1.5.Metodologi Penelitian

Metode penelitian merupakan cara utama yang digunakan peneliti untuk mencapai tujuan dan menentukan jawaban atas masalah yang diajukan. Pada penelitian kali ini penulis akan menggunakan metode penelitian deskriptif.

Metode penelitian deskriptif adalah suatu metode dimana setiap objek digambarkan secara jelas dan nyata sesuai dengan fakta. Metode yang digunakan selama penelitian di gambarkan pada gambar 1.1. di bawah:



Gambar 1.1 Metode Penelitian

Penjelasan metode penelitian yang digunakan pada gambar 1.1 adalah sebagai berikut :

a. Identifikasi Masalah

   Berdasarkan latar belakang yang telah disebutkan, permasalahan yang dapat di identifikasi adalah kebutuhan masyarakat akan peringkas teks otomatis yang dapat menghasilkan ringkasan dengan cepat dan akurat. Serta penggunaan metode pada peringkas teks otomatis yang dapat menghasilkan ringkasan dengan tingkat akurasi tinggi.

b. Analisis Masalah

   Dari permasalahan yang telah di identifikasi didapatkan bahwa salah satu metode yang digunakan untuk menghasilkan ringkasan pada peringkas teks

otomatis untuk artikel berita berbahasa Indonesia adalah metode *Latent Semantic Analysis* (LSA). Tetapi metode LSA yang diterapkan pada bebrapa penelitian untuk menghasilkan ringkasan pada dokumen berbahasa Indonesia menggunakan metode LSA yang lama sedangkan metode LSA telah mengalami beberapa tahap pengembangan. Sehingga ringkasan yang dihasilkan oleh metode LSA yang lama memiliki tingkat akurasi yang lebih rendah jika dibandingkan dengan metode LSA yang telah dikembangkan.

c. Pengambilan Hipotesis Awal

   Dari hasil analisis masalah di atas dapat diambil sebuah hipotesis awal yaitu, untuk menghasilkan sebuah peringkas teks otomatis metode yang dapat digunakan adalah *Cross Method Latent Semantic Analysis* yang dapat menghasilkan sebuah ringkasan yang lebih akurat dibandingkan dengan metode *latent semantic analysis* biasa.

d. Studi Literatur

   Pada tahap ini dilakukan studi terhadap literatur – literatur yang berkaitan dengan penelitian yang dilakukan seperti literature mengenai peringkas teks otomatis, proses pengolahan teks, dan *cross method latent semantic analysis*, serta literature – literatur lain yang mendukung penelitian.

e. Perancangan Simulator

   Pada tahap ini dilakukan proses perancangan peringkas teks otomatis. Proses perancangan terdiri dari proses analisis kebutuhan fungsional dan non fungsional dari peringkas teks otomatis yang akan dibangun, analisis data masukan, perancangan tampilan antar muka peringkas teks otomatis dan perancangan jaringan semantik.

f. Pembangunan Peringkas Teks Otomatis

   Pada tahap ini dilakaukan pembangunan dari peringkas teks otomatis untuk teks berbahasa Indonesia. Dimulai dari tahap pembangunan *interface* peringkas teks otomatis, tahapan *preprocessing*(pemecahan kalimat, *Case Folding, Tokenizing, Stop Words Removal* dan *Stemming*), pembobotan kata menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-

IDF) dan implementasi C*ross Method Latent Semantic Analysis* untuk menghasilkan ringkasan.

g. Pengujian Peringkas Teks Otomatis

Pada tahapan ini, akan dilakukan proses pengujian terhadap peringkas teks otomatis yang telah di bangun untuk melihat apakah peringkas teks otomatis dapat berjalan dengan semestinya. Jika peringkas teks otomatis yang dibangun memiliki kekurangan maka akan kembali ke tahapan perancangan peringkas teks otomatis untuk melakukan analisis kesalahan yang mungkin terjadi pada saat pembangunan peringkas teks otomatis. Jika peringkas teks otomatis sudah dapat berjalan dengan semestinya maka akan masuk ke dalam tahapan pengujian metode.

h. Pengujian Metode

Dalam tahapan ini akan dilakukan pengujian terhadap *Cross Method Latent Semantic Analysis* untuk mengetahui hasil dari implementasi metode ini dalam peringkas teks otomatis. Jika dari tahapan ini memberikan hasil yang kurang baik maka akan kembali ke tahapan studi literatur untuk menganalisis kesalahan yang mungkin terjadi pada saat penerapan metode.

i. Analisis Hasil Ringkasan

Pada proses ini dilakukan analisis pada hasil ringkasan yang didapatkan dengan menggunakan metode *Latent Semantic Analysis*. Untuk proses analisis digunakan metode *Precision,Recall* serta metode *F-Measure* untuk mengetahui tingkat ke akuratan ringkasan yang dihasilkan dari *Cross Method Latent Semantic Analysis.*

j. Pengambilan Kesimpulan

Pada tahap ini dilakukan pengambilan kesimpulan yang didapat dari hasil ringkasan yang dihasilkan oleh peringkas teks otomatis untuk teks berita berbahasa Indonesia.

1.6. Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mengenai permasalahan dan pemecahannya. Sistematika penulisan tugas akhir ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini membahas mengenai latar belakang , perumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian, serta sistematika penulisan untuk menjelaskan pokok – pokok pembahasannya.

BAB 2 LANDASAN TEORI

Pada bab ini akan menjelaskan mengenai teori – teori pendukung yang berhubungan dengan penelitian dan pembangunan peringkas teks otomatis.

BAB 3 ANALISIS KEBUTUHAN ALGORITMA

Dalam bab ini berisikan analisis dan perancangan dari peringkas teks otomatis yang akan dibangun. Proses analisis meliputi proses analisis masalah, analisis data masukan, analisis kebutuhan dari perangkat lunak yang akan dibangun,analisis tahap *preprocessing* dan analisis dari metode peringkasan yang digunakan dalam penelitian. Sedangkan pada proses perancangan peringkas teks otomatis akan di bahas mengenai proses perancangan basis data,tampilan antar muka dan jaringan semantic yang akan diterapkan pada peringkas teks otomatis.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Dalam bab ini berisikan implementasi dari sistem yang telah dirancang untuk mengetahui keberhasilan dari pembangunan peringkas teks otomatis serta tingkat keberhasilan algoritma dalam menghasilkan sebuah ringkasan pada artikel berita berbahasa Indonesia. Pengujian pada peringkas teks otomatis dilakukan dengan menggunakan metode pengujian *Precision, Recall* dan *F-Measure* yang digunakan untuk mengukur tingkat akurasi dari ringkasan yang dihasilkan.

BAB 5 KESIMPULAN DAN SARAN

Bab ini bereisi mengenai kesimpulan yang diperoleh dari hasil pengujian implementasi *Cross Method Latent Semantic Analysis* pada peringkas teks otomatis untuk menghasilkan ringkasan serta saran untuk pengembangan peringkas teks otomatis kedepan.