BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Sistem Temu Kembali (*Information Retrieval System*) dirancang untuk menemukan data (*file*) atau informasi yang diperlukan. Sistem Temu Kembali (*Information Retrieval System*) bertujuan untuk menjembatani kebutuhan informasi dengan sumber informasi yang tersedia secara relevan. *Generalized Vector Space Model* (GVSM) merupakan salah satu model Sistem Temu Kembali (*Information Retrieval System*).

Dalam hal ini, sistem temu kembali informasi berkaitan dengan representasi, penyimpanan, dan akses terhadap representasi *file*. *File* yang ditemukan harus relevan dengan kebutuhan informasi yang dinyatakan dalam *query*. Terkadang apa yang dicari tidak relevan dengan apa yang diinginkan, terdapat beberapa penyebab yang mengakibatkan apa yang dicari tidak relevan dengan apa yang diinginkan, salah satunya adalah kata kunci (*keyword*) yang ambigu[13]. Solusi yang dapat digunakan untuk mengoptimalkan kata kunci (*keyword*) adalah dengan mengimplementasikan algoritma lesk. Algoritma lesk merupakan algoritma yang digunakan untuk menemukan makna kata yang berkaitan dengan kata kunci utama, dengan kata lain terdapat kata yang tidak dimasukan kedalam *query* tapi masuk kedalam kata pencarian.

Berdasarkan latar belakang yang telah diuraikan, tujuan dari tugas akhir ini yaitu untuk menilai efektifitas algoritma lesk dalam melakukan pencarian kata atau kalimat yang ambigu. Penelitian sebagain topik untuk tugas akhir ini mengangkat judul **"IMPLEMENTASI METODE GENERALIZED VECTOR SPACE MODEL (GVSM) MENGGUNAKAN ALGORITMA LESK PADA SISTEM TEMU KEMBALI"**.

1.2. Perumusan Masalah

Berdasarkan penjelasan yang dipaparkan dalam latar belakang sebelumnya, maka rumusan dalam masalah ini adalah menghitung efektifitas algoritma lesk dalam melakukan pencarian pada sekumpulan data (*file*) di dalam perangkat komputer berdasarkan kata atau kalimat menggunakan metode *Generalized Vector Space Model* (GSVM) pada Sistem Temu Kembali (*Information Retrieval System*).

1.3. Maksud dan Tujuan

Berdasarkan masalah yang telah diuraikan pada bagian latar belakang dan perumusan masalah, maka maksud dari penelitian ini adalah melakukan implementasi algoritma lesk pada metode *Generalized Vector Space Model* (GVSM).

Tujuan dari penelitian ini adalah untuk menilai efektifitas *recall, precission*, dan bobot similiaritas pada algoritma lesk dalam melakukan pencarian kata atau kalimat yang ambigu di dalam data (*file*) atau informasi menggunakan metode *Generalized Vector Space Model* (GVSM).

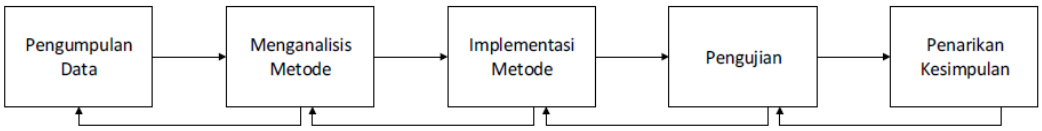
1.4. Batasan Masalah

Aplikasi yang dibangun adalah aplikasi dengan batasan masalah sebagai berikut:

1. Dokumen yang dicari menggunakan format *.doc
2. Data masukan (*query*) menggunakan bahasa Indonesia dengan Ejaan Yang Disempurnakan (EYD) berdasarkan Keputusan Menteri Pendidikan dan Kebudayaan tanggal 27 Agustus 1975 Nomor 0196/U/1975[5].
3. Isi dokumen menggunakan bahasa Indonesia.
4. Gambar dan diagram yang terdapat di dokumen tidak dapat terbaca.
5. Dokumen yang digunakan berasal dari laporan penelitian dan laporan kerja praktek di perpustakaan UNIKOM.
6. Kamus sinonim dan makna kata diambil secara manual di Kamus Besar Bahasa Indonesia (KBBI).

1.5. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini menggunakan beberapa tahap, yaitu tahap pengumpulan data, analisis metode, implementasi metode, pengujian dan penarikan kesimpulan yang dapat dilihat pada Gambar 1.1 berikut ini:



Gambar 1.1 Metodologi Penelitian

1.5.1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan adalah studi literatur, yaitu pengumpulan data yang bersumber dari buku, jurnal, paper dan situs internet yang berhubungan dengan Sistem Temu Kembali (*Information Retrieval System*), *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word removal*, *stemming*, dan algoritma lesk untuk dijadikan referensi penulisan dan penelitian.

1.5.2. Menganalisis Metode

Pada tahapan ini dilakukan analisis metode yang digunakan yaitu Sistem Temu Kembali *(Information Retrieval System*), *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word removal, stemming* dan algoritma lesk.

1.5.3. Implementasi Metode

Setelah melakukan analisis metode maka pada tahap ini dilakukan implementasi metode tersebut yang dibangun menggunakan bahasa C# dengan *database* SQL Server sesuai perancangan yang dibuat.

1.5.4. Pengujian

Setelah aplikasi pencarian informasi kata atau kalimat pada kumpulan dokumen yang dibangun maka tahap selanjutnya adalah melakukan pengujian *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word*

removal, *stemming* dan algoritma lesk. Pada tahapan ini akan dilihat keefektifan aplikasi dalam melakukan pencarian informasi.

1.5.5. Penarikan Kesimpulan

Langkah selanjutnya adalah melakukan penarikan kesimpulan berdasarkan pengujian yang telah dilakukan sebelumnya. Penarikan kesimpulan dari hasil penerapan metode *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word removal*, *stemming* dan algoritma lesk.

1.6. Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang masalah, perumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang urutan pemahaman dalam menyajikan penelitian ini.

BAB II LANDASAN TEORI

Bab ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian yang dilakukan dan hal-hal yang berguna dalam proses analisis permasalahan serta tinjauan terhadap penelitian-penelitian serupa yang telah pernah dilakukan sebelumnya yaitu mengenai Sistem Temu Kembali *(Information Retrieval System*), *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word removal*, *stemming* dan algoritma lesk.

BAB III ANALISIS DAN KEBUTUHAN ALGORITMA

Bab ini berisi tentang analisis perhitungan metode *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word removal*, *stemming* dan algoritma lesk dalam melakukan penelitian.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang implementasi dari hasil analisis dan perancangan yang telah dibuat serta pengujian metode *Generalized Vector Space Model* (GVSM), *tokenizing*, *filtration* atau *stop word removal*, *stemming* dan algoritma lesk.

BAB V KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil penelitian serta saran-saran untuk pengembangan kedepannya.