

## **BAB I**

### **PENDAHULUAN**

#### **I.1. Latar Belakang Masalah**

Dalam era teknologi seperti saat ini, informasi berupa teks sudah tidak lagi selalu tersimpan dalam media cetak seperti kertas. Orang sudah mulai cenderung menyimpan informasi secara digital. Karena lebih mudah dalam penyimpanan dan cepat. Tuntutan dari gerakan anti *global warming* juga mendukung penyimpanan informasi secara digital, baik untuk keperluan pribadi, perkantoran, atau bahkan pemerintah. Selain itu, informasi sudah menjadi suatu komoditas yang dapat diperjualbelikan. Orang sampai rela membayar mahal, baik dalam waktu maupun uang demi mendapatkan informasi, keakuratan dari informasi yang mereka butuhkan, supaya mereka mendapatkan dengan cepat. Tidak hanya kecepatan mendapatkan informasi, keakuratan dari informasi juga menjadi harapan semua orang.

Banyaknya dokumen berformat teks seperti jurnal, buku, dan berita yang sudah tersimpan secara digital, muncul permasalahan dimana informasi yang tadinya tersedia dengan baik menjadi kabur/hilang karena terlalu banyak dokumen yang tersimpan dalam media penyimpanan digital. Imbasnya proses mencari informasi tertentu yang dibutuhkan dari dokumen-dokumen tersebut menjadi makin sulit dan lama.

Masalah lain akan terjadi ketika setiap dokumen tersebut ingin dikategorikan kedalam kelas-kelas tertentu, karena harus dilihat, dibaca, dan dipahami isi tiap dokumen dalam data. Setelah selesai membaca seluruh isi data, barulah bisa ditentukan kelas-kelas bagi dokumen dan membagi dokumen dalam kelas tersebut.

Penelitian ini menggunakan teknik *text mining* dengan algoritma *vector space model* untuk pengukuran kemiripan hasil pencarian terhadap dokumen. Teknik *text mining* terdapat tiga tahapan penting yaitu: *text preprocessing*, *text transformation*, dan *pattern discovery*. Tahapan *text preprocessing* terdiri tahapan pembersihan teks dan pemecahan kalimat menjadi kata-kata (*tokenizing*). Tahapan *text transformation* terdiri dari tahapan *filtering* dan *stemming*. Karena tahapan *stemming* lebih kompleks akan dirubah ke tahap perbandingan kata. Tahapan *pattern discovery* adalah tahapan pembobotan, pembobotan yang digunakan yaitu pembobotan *TF-IDF* dan *vector space model* dengan metode *cosine similarity*.

## **I.2. Rumusan Masalah**

Berdasarkan latar belakang diatas maka diperoleh beberapa rumusan masalah yang akan dibahas pada pembuatan tugas akhir ini adalah:

- a. Bagaimana implementasi proses pencarian dokumen berformat teks yang dilakukan oleh metode *cosine similarity* yang dipadukan dengan pembobotan *TF-IDF*.
- b. Bagaimana mengukur akurasi pencarian dokumen dengan metode *cosine similarity*.

### **I.3. Tujuan**

Tujuan dari pembuatan tugas akhir ini adalah:

- a. Membangun perangkat lunak pencarian dokumen untuk membantu penemuan kembali informasi yang hilang karena data yang banyak.
- b. Menguji keakuratan metode *cosine similarity* dengan pembobotan *TF-IDF* dalam pencarian dokumen teks.

### **I.4. Batasan Masalah**

Dalam melakukan pembuatan tugas akhir ini, peneliti membatasi masalah atau ruang lingkup penelitian. Hal ini dimaksudkan agar penelitian dapat dilakukan pada batasan yang jelas.

Batasan masalah dalam penelitian ini adalah:

- a. Dokumen teks uji dalam penelitian ini terbatas hanya berekstensi .txt (plan teks).
- b. Dokumen teks uji adalah teks berbahasa Indonesia.
- c. Aplikasi dibangun berbasis web dengan menggunakan bahasa pemrograman C#.
- d. *Database server* menggunakan SQL Server 2005.

### **I.5. Metode Penelitian**

- a. Studi Literatur

Metode ini dilakukan karena untuk mendapatkan hasil dari penelitian, maka dibutuhkan hasil penelitian sebelumnya sebagai acuan yang akan dilakukan.

- b. Pembangunan Perangkat Lunak

Metode ini dilakukan dengan mengembangkan perangkat lunak yang sesuai dengan analisis kebutuhan. Metode ini terbagi menjadi empat tahap, yaitu:

1. Analisis Kebutuhan Perangkat Lunak

Pada tahap ini yang dilakukan adalah menganalisis kebutuhan perangkat lunak.

2. Perancangan Perangkat Lunak

Adalah tahap untuk membangun rancangan basis data relasional, rancangan antarmuka, dan rancangan implementasi perangkat lunak.

3. Pengkodean Perangkat Lunak

Adalah tahap untuk mengimplementasikan perangkat lunak dari hasil analisis dan perancangan dalam suatu bahasa pemrograman.

4. Pengujian Perangkat Lunak

Adalah tahap untuk melakukan pengujian terhadap kebutuhan dan fungsionalitas perangkat lunak yang telah dibangun.

#### **I.6. Sistematika Penulisan**

Laporan ditulis berdasarkan tata tulis laporan yang telah ditetapkan oleh Program Studi Teknik Informatika Universitas Atma Jaya Yogyakarta, sebagai berikut:

#### **BAB I : Pendahuluan**

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penyusunan, batasan masalah, cara penelitian, dan sistematika penulisan Laporan Tugas Akhir.

**BAB II : Tinjauan Pustaka**

Bab ini berisi jurnal-jurnal ilmiah yang berkaitan dengan topik yang dibahas, dan tabel perbandingan antara sistem yang telah ada dengan sistem yang akan dibuat.

**BAB III : Landasan Teori**

Bab ini berisi penjelasan mengenai beberapa teori yang relevan dengan permasalahan yang dibahas dalam penulisan Laporan Tugas Akhir.

**BAB IV : Analisis dan Perancangan Sistem**

Bab ini menganalisis permasalahan yang diatasi dengan membangun model serta membahas perancangan perangkat lunak.

**BAB V : Implementasi Pengujian Sistem, Perhitungan dengan Algoritma *TF-IDF* dan *VSM***

Bab ini berisi mengenai implementasi perangkat lunak yang telah dibuat, gambaran umum sistem, pengujian sistem, dan perhitungan data-data dengan algoritma *TF-IDF* dan *VSM*.

**BAB VI : Kesimpulan dan Saran**

Bab ini berisi kesimpulan-kesimpulan dari pembahasan tugas akhir secara keseluruhan dan saran untuk pengembangan lebih lanjut.

**DAFTAR PUSTAKA**

**LAMPIRAN**