BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

PT. eBdesk Indonesia adalah sebuah produsen piranti lunak yang telah berdiri sejak tahun 2000. Saat ini PT eBdesk sedang mengembangkan sebuah produk bernama Collaboration, yaitu berupa sebuah aplikasi yang memungkinkan penggunanya berkolaborasi untuk mencapai suatu tujuan. Beberapa fitur utama dalam aplikasi Collaboration adalah *sharing* dokumen, galeri, blog dan forum.

Dalam aplikasi Collaboration ini terdapat fasilitas untuk menyimpan dokumen. Proses pencarian dokumen yang dilakukan pada aplikasi collaboration saat ini baru sebatas mencari dokumen berdasarkan judul dan deskripsi singkat yang disimpan ke dalam *database*.

Untuk lebih menambah nilai jual dari aplikasi collaboration ini, oleh karena itu ingin dibuatkan sebuah fitur baru yaitu pencarian dokumen yang bisa menelusuri *query* pencarian sampai ke dalam isi dokumen yang di-*upload* nya itu sendiri.

Proses dalam pencarian dokumen dapat digambarkan sebagai sebuah proses untuk mendapatkan dokumen yang relevan dari beberapa dokumen yang ada pada koleksi dokumen melalui pencarian *query* yang dimasukkan pengguna. Masalah yang sering dihadapi dalam membuat aplikasi pencarian adalah bagaimana menyajikan dokumen yang relevan dengan kebutuhan pengguna dari sekian banyak dokumen yang tersedia.

Untuk menghasilkan dokumen yang relevan terhadap *query* pencarian maka diperlukan proses perangkingan pada dokumen hasil pencarian. Dengan melakukan perangkingan terhadap dokumen, dokumen yang telah didapatkan tersebut ditampilkan terurut dari dokumen yang memiliki tingkat relevansi lebih tinggi ke tingkat relevansi rendah.

Ada beberapa model yang digunakan dalam membuat aplikasi pencarian dokumen yaitu : Model Boolean, Model Ruang Vektor (*Vector Space Model*), dan Model Probabilistik. Namun dari ketiga model tersebut model ruang vektor adalah model yang paling popular dan umum digunakan karena mudah diimplementasikan dan mendukung adanya pemeringkatan / perangkingan dokumen.

Dalam model ruang vektor, metode yang digunakan untuk melakukan perangkingan adalah dengan menghitung bobot dokumen terhadap *query*. Pembobotan ini menghitung bobot frekuensi kemunculan kata dalam suatu dokumen (*term frequency – tf*) diimbangi dengan frekuensi kemunculan kata pada koleksi dokumen(*inverse document frequency - idf*), yang dikenal dengan pembobotan tf-idf. Pembobotan menggunakan tf-idf menunjukkan bahwa deskripsi terbaik dari dokumen terjadi ketika kata yang banyak muncul dalam dokumen tersebut dan sangat sedikit muncul pada dokumen yang lain.

Tf-idf adalah algoritma yang efisien dan sederhana untuk pencocokan kata dalam *query* untuk dokumen yang relevan terhadap *query* tersebut. Tf-idf mengembalikan dokumen yang sangat relevan terhadap *query*. (Ramos, 2002)

Berdasarkan uraian diatas, penulis ingin melakukan implementasi tf-idf dengan dengan membuat perangkat lunak. Perangkat lunak ini diharapkan bisa melakukan pembobotan atau perangkingan pada pencarian dokumen yang ada.

1.2 Identifikasi Masalah

Permasalahan yang akan diangkat dalam skripsi ini adalah :

1. Bagaimana membuat aplikasi pencarian dokumen yang bisa menelusuri *query* ke dalam isi teks dokumen ?
2. Bagaimana mengimplementasikan metode tf-idf pada aplikasi pencarian dokumen tersebut ?

1.3 Maksud dan Tujuan

Maksud dari pembuatan skripsi ini yaitu untuk membuat sebuah aplikasi pencarian dokumen yang mengimplementasikan penggunaan metode tf-idf.

Tujuan yang akan dicapai dalam pembuatan skripsi ini adalah :

1. Membuat aplikasi yang bisa mencari dokumen dari koleksi dokumen berdasarkan *query* yang diinputkan pengguna, dan memberikan peringkat terhadap hasil pencarian tersebut supaya menghasilkan dokumen yang relevan terhadap *query*.
2. Membuat aplikasi pencarian dokumen berbasis web yang mengimplementasikan metode tf-idf dan akan menampilkan pencarian dokumen dari seluruh dokumen yang ada dalam aplikasi.

1.4 Metodologi Penelitian

Tahapan-tahapan yang akan dilalui dalam pelaksanaan skripsi ini adalah sebagai berikut :

a. Studi literatur

   Mengumpulkan bahan-bahan referensi yang menunjang proses penelitian yaitu yang berhubungan dengan pengenalan *search engine*, penentuan indeks dan pencarian pada *search engine* dan pembobotan pada *search engine*.

b. Pengumpulan data di lapangan

   Pengumpulan data ini dilakukan dengan dua cara, pertama dengan observasi langsung ketempat penelitian untuk mendapatkan data primer dan dokumen-dokumen yang diperlukan yang berkaitan dengan penelitian. Kedua, melakukan wawancara mendalam. Wawancara dilakukan untuk memperoleh data-data yang diperlukan penulis. Wawancara ini juga bertujuan untuk mencari tahu kondisi aplikasi yang saat ini berjalan di lapangan, dan masalah apa yang dihadapi.

c. Analisis

   Melakukan analisis terhadap sistem yang berjalan saat ini, menganalisis konsep pada *text mining*, serta menganalisis metode atau algoritma yang akan digunakan untuk menyelesaikan permasalahan, termasuk menentukan bahasa pemrograman yang digunakan.

d. Perancangan

Pada tahap ini dilakukan perancangan terhadap sistem yang akan dibangun. Diantaranya adalah : membuat rancangan database, rancangan fungsionalitas, dan membuat rancangan antarmuka sistem.

e. Implementasi

Pada tahap ini dilakukan implementasi berdasarkan hasil rancangan. Adapun proses yang dilakukan pada tahap implementasi antara lain :

1. Pembangunan database berbasis MySQL.
2. Perancangan *interface.*
3. Pengimplementasian algoritma metode TF-IDF yang diintegrasikan dengan studi kasus.

f. Pengambilan kesimpulan dan saran

Pada tahap ini adalah tahap terakhir dari penelitian yang dilakukan yakni pengambilan kesimpulan dari beberapa tahap yang telah dilalui. Kemudian menyusun laporan dari penelitian yang dilakukan.

1.5 Batasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang diketahui, penulis membatasi ruang lingkupnya sebagai berikut :

1. Membuat aplikasi pencarian dokumen yang diletakkan di dalam aplikasi eBdesk Collaboration.

2. Menambahkan proses penentuan indeks pada aplikasi upload dokumen dalam eBdesk Collaboration, untuk mendukung proses pencarian dokumen.
3. Pencarian dalam aplikasi ini berupa teks berbahasa Inggris dan pengujian dilakukan terhadap dokumen teks yang berisi kata (*term*). Dikarenakan aplikasi eBdesk Collaboration saat ini diimplementasikan pada proyek di Malaysia, dan diharapkan akan bisa diimplementasikan di negara tetangga lainnya seperti Singapura.
4. Pengideksan dan pencarian dilakukan terhadap dokumen yang berekstensi *.txt dan *.doc.
5. Ukuran *file* yang di-*upload* maksimal 4 MB.

1.6 Sistematika Penulisan

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

BAB 1 PENDAHULUAN

Berisi tentang beberapa hal umum tentang maksud dan tujuan penulisan skripsi serta pembelajaran terhadap metode tf-idf sebagai acuan dalam pembuatan aplikasi pencarian dokumen, yang terdiri dari latar belakang, identifikasi masalah, batasan masalah, maksud dan tujuan, dan metode pendekatan.

BAB 2 TINJAUAN PUSTAKA

Menjelaskan tentang konsep dasar aplikasi *text mining*, penentuan indeks (*tokenizing*, *filtering*, *stemming*, *tagging*), metode tf-idf, *vector space model*, perancangan sistem, basis data, tinjauan perangkat lunak, dan bahasa pemrograman yang digunakan untuk membangun aplikasi.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Membahas analisis terhadap kebutuhan aplikasi dengan menggunakan metode tf-idf, perancangan basis data, perancangan program aplikasi, perancangan *input* dan *output*.

BAB 4 IMPLEMENTASI

Dalam bab ini membahas dan menampilkan aplikasi yang sudah dibangun berdasarkan perancangan yang telah dibuat dan menunjukkan hasil dari implementasi aplikasi terhadap data yang diolah. Aplikasi dibuat menggunakan bahasa pemrograman Java dan basis data MySQL.

BAB 5 KESIMPULAN DAN SARAN

Menyampaikan saran dari penulis yang dianggap perlu untuk pengembangan aplikasi dalam skripsi ini dimasa yang akan datang, dan memaparkan kesimpulan dari materi skripsi yang dibahas.