

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Informasi telah menjadi kebutuhan primer pada kehidupan saat ini. Pesatnya perkembangan teknologi dewasa ini telah menyebabkan aliran informasi begitu lancar sehingga terwujud kemudahan untuk memperoleh informasi, salah satunya melalui internet. Saat ini, informasi yang tersedia di internet semakin banyak dan terus meningkat dengan tajam. Informasi-informasi tersebut tersedia dalam berbagai format, seperti teks, *audio*, dan *visual*. Dengan semakin banyak dan beragamnya informasi yang tersedia, kebutuhan pengguna internet telah bergeser dari arah kuantitatif ke arah kualitatif. Kebutuhan yang semula berupa informasi sebanyak-banyaknya telah bergeser menjadi informasi yang secukupnya asalkan relevan dengan keperluan (Jaya, 2007).

Salah satu cara untuk mendapatkan informasi yang akurat dan relevan adalah dengan menggunakan sistem temu balik informasi, “sistem ini membuat perhitungan untuk menentukan apakah sebuah informasi relevan dengan kebutuhan penggunanya” (Ramadhany, 2008). Salah satu aplikasi umum dari sistem temu balik informasi adalah *search engine* atau mesin pencari yang terdapat pada jaringan internet. Pengguna dapat mencari halaman-halaman web yang dibutuhkan melalui *search engine*. Walaupun tersedia secara gratis dan dalam jumlah banyak, keberadaan *search engine* di internet dirasakan masih kurang dari aspek kualitas. Hal ini wajar, mengingat hasil pencarian yang diberikan oleh mesin-mesin pencari tersebut seringkali dalam jumlah besar namun

kurang relevan. Oleh karena itu, kebutuhan akan suatu mekanisme pencarian dokumen yang lebih efektif dirasakan semakin mendesak (Mandala dan Setiawan, 2002; Jaya, 2007).

Di dalam bidang sistem temu balik informasi, dikenal berbagai model untuk menilai secara objektif presisi dari suatu pencarian, antara lain model *Boolean* (*Boolean Model*), model ruang vektor (*Vector Space Model*) dan model probabilistik (*Probabilistic Model*). Ketiga model ini masing-masing memiliki kekurangan dan kelebihan. Model *Boolean* adalah model yang paling awal dikenal dan paling mudah untuk diimplementasikan. Hanya saja, model *Boolean* tidak dapat memberikan hasil yang diharapkan dan sangat lambat dalam *run-time* (Jaya, 2007).

Sistem temu balik informasi juga memperkenalkan model probalistik. Model ini mengurutkan dokumen dalam urutan menurun terhadap peluang relevansi sebuah dokumen pada informasi yang dibutuhkan (Ramadhany, 2008).

Model berikutnya adalah model ruang vektor yang saat ini sangat populer dalam sistem temu balik informasi. Model ini berhasil memberikan hasil yang lebih baik dibandingkan model *Boolean*. Model ini juga dapat menampilkan hasil temu balik secara terurut (*ranking*) (Jaya, 2007). Model ruang vektor tidak membutuhkan komputasi yang berlebihan sehingga waktu untuk mengeksekusi akan semakin cepat dan lebih efektif (Ramadhany, 2008). Hal inilah yang menjadi alasan penggunaan model ruang vektor di dalam skripsi ini.

Saat ini, telah ada *open source library* sistem temu balik informasi yaitu *Apache Lucene* yang dibuat di atas *platform* Java. Akan tetapi, sangat sulit untuk

mencari hostingan Java dan harganya pun mahal. Oleh karena itu, penulis membuat sistem temu balik informasi dengan *platform* PHP, karena untuk mendapatkan hostingan PHP itu sangat mudah dan telah banyak tersedia.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, terdapat beberapa permasalahan yang dapat dirumuskan, antara lain:

1. Bagaimana membangun aplikasi sistem temu balik informasi dengan menggunakan model ruang vektor dalam suatu perangkat lunak berbasis web berupa mesin pencari?
2. Bagaimana mesin pencari dapat menghasilkan dokumen yang relevan dan terurut berdasarkan tingkat korelevanannya dengan *query* yang dimasukkan oleh pengguna?

1.3 Batasan Masalah

Dalam pengerjaan skripsi ini, ditetapkan beberapa batasan masalah, antara lain sebagai berikut:

1. Model yang digunakan dalam sistem temu balik informasi ini adalah model ruang vektor.
2. Perangkat lunak yang dibangun berupa *prototype*.
3. Dokumen yang digunakan adalah dokumen berita berbahasa Indonesia.

1.4 Tujuan Penelitian

Sejalan dengan permasalahan yang telah dirumuskan, maka tujuan yang ingin dicapai pada penelitian ini adalah:

1. Membangun aplikasi sistem temu balik informasi dengan menggunakan model ruang vektor dalam suatu perangkat lunak berbasis web berupa mesin pencari.
2. Menghasilkan dokumen yang relevan dan terurut berdasarkan tingkat korelevanannya dengan *query* yang dimasukkan oleh pengguna.

1.5 Manfaat Penelitian

1. Aplikasi ini dapat menjadi solusi untuk memperoleh informasi yang sesuai dengan pengguna.
2. Memberikan informasi kepada khalayak tentang penggunaan model ruang vektor dalam sistem temu balik informasi.

1.6 Metode Penelitian

Metode yang diterapkan dalam pembuatan skripsi ini, antara lain:

1. Eksplorasi dan Studi Literatur

Eksplorasi dan studi literatur dilakukan dengan mempelajari konsep-konsep yang berkaitan dengan skripsi ini, seperti sistem temu balik informasi, model ruang vektor, *text preprocessing*, *indexing*, dan pembobotan TF-IDF melalui literatur-literatur seperti buku, paper, dan sumber ilmiah lain seperti situs internet ataupun artikel dokumen teks yang berhubungan.

2. Analisis dan Perancangan Perangkat Lunak

Analisis dan perancangan perangkat lunak dilakukan untuk menentukan permasalahan mengenai bahasa pemrograman yang akan

digunakan, struktur data, *input/output* program, dan permasalahan teknik algoritma yang akan diimplementasikan.

3. Implementasi Program

Detail mengenai implementasi program dilakukan sesuai hasil analisis pada tahapan sebelumnya.

4. Hasil Akhir dan Penarikan Kesimpulan

Analisis hasil dilakukan untuk mengetahui performansi model ruang vektor dalam sistem temu balik informasi pencarian dokumen, jika ternyata hasilnya baik, maka dilakukan analisis akhir untuk mengetahui penyebabnya, selanjutnya dilakukan penarikan kesimpulan.

1.7 Sistematika Penulisan

Dalam penyusunan skripsi ini, sistematika penulisan dibagi menjadi beberapa bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas masalah yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bagian ini memuat landasan teori yang berfungsi sebagai sumber atau alat dalam memahami permasalahan yang berkaitan dengan teori dan konsep sistem temu balik informasi, model ruang vektor, *text preprocessing*, *indexing*, dan pembobotan TF-IDF.

BAB III METODOLOGI PENELITIAN

Bab ini merupakan penjabaran dari alat dan bahan penelitian, metode pengembangan sistem yang digunakan yaitu pendekatan terstruktur dengan model prosesnya sekuensial linier.

BAB IV HASIL PENELITIAN

Bab ini akan membahas secara mendalam hal-hal yang menjadi rumusan masalah untuk mencari solusi atas rumusan masalah tersebut.

BAB V PENUTUP

Bab ini akan memberikan kesimpulan yang merupakan jawaban atas rumusan masalah dalam penelitian, serta memberikan saran yang dapat dijadikan rekomendasi dalam melakukan penelitian selanjutnya.

