BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Saat ini informasi digital berkembang sangat pesat. Setiap hari jutaan informasi baru terus bertambah. Berdasarkan survey IDC *(Internasional Data Corporation)* tahun 2007 data digital yang telah dibuat, ditangkap, dan direplikasi oleh seluruh penduduk dunia adalah sebesar 281 miliar giga bytes (281 exa bytes). Data ini terdiri dari dokumen dalam berbagai macam format seperti gambar, email, video, web page, teks, dll. 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.

Seiring dengan bertambahnya hal tersebut, kebutuhan di dalam mencari informasi elektronis juga semakin meningkat. Dalam hal ini Robert R.K. mengindikasikan timbulnya dua masalah besar, yaitu teknologi penyimpanan informasi (*Information Storage)* dan temu kembali informasi (*Information Retreival).* Oleh karena itu, Saat ini telah banyak dari berbagai informasi tersebut dapat diakses secara elektronik melalui *internet* dengan menggunakan berbagai mesin pencari (*search engine).* Perbedaan mesin penelusur yang satu dengan yang lain

sangat bergantung pada teknik temu kembali informasi dan teknik pengindeksan yang dipakai.

Dari hasil pengamatan langsung terhadap portal website berita berbahasa Indonesia terutama dalam berita IT, karena sedikit sekali website portal berita IT yang menyediakan mesin pencari bagi website itu sendiri. Kebanyakan dari website-website tersebut menyediakan mesin pencari yang langsung terhubung ke mesin pencari google, sehingga proses pengindeksan kata kunci dengan dokumen tidak akan langsung dihubungkan dengan konten berita yang ada dalam website tersebut. Maka dari hasil pencarian tersebut ada yang relevan dan ada yang kurang relevan dengan kata kunci.

Berbagai macam penelitian di area ini telah banyak dilakukan, salah satunya adalah melalui disiplin ilmu IR (*Information Retrieval).* IR merupakan bidang yang mengkaji metode-metode di dalam pencarian dokumen berdasarkan representasi kebutuhan informasi berupa kata kunci, yaitu *keyword* atau *query*.

Dalam *Information Retrieval System,* Salton menjelaskan bahwa terdapat 3 model yang digunakan, yaitu *Boolean Model, Probabilistic Model,* dan *Vector Space Model.* Model terakhir inilah yang paling sederhana dan paling produktif. Model ruang vektor ini merepresentasikan *term* yang digunakan baik oleh dokumen maupun oleh *query.* Elemen *vektor* tersebut merupakan bobot *term* yang menerepkan

pembobotan kombinasi berupa perkalian bobot lokal *term frequency (tf)* dan global *inverse document frequency (idf).*[5]

Berdasarkan permasalahan diatas akan dilakukan suatu analisis terhadap *Information Retrieval System* menggunakan Model Ruang Vektor.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang masalah yang dikemukakan, maka dapat dirumuskan beberapa masalahnya sebagai berikut :

1. Bagaimana proses sistem temu balik informasi atau *Information Retrieval System.*
2. Bagaimana Implementasi *Information Retrieval System* pada mesin pencari berita IT.
3. Bagaimana pencarian dokumen yang relevan di *Information Retrieval System* dengan menggunakan Model Ruang Vektor.

1.3 Maksud dan Tujuan

Maksud dari ditulisnya skripsi ini adalah untuk melakukan analisis dan mengimplementasikan *Information Retrieval System* pada web menggunakan *vector space model* (model ruang vektor).

Adapun tujuan yang akan dicapai dari penulisan skripsi ini adalah sebagai berikut :

1. Menganalisis proses *Information Retrieval System* dalam melakukan pencarian dokumen berita.
2. Mengimplementasikan *Information Retrieval System* pada *website* mesin pencari berita IT.
3. Menerapkan Model Ruang Vektor atau *Vektor Space Model (VSM)* pada *Information Retrieval System.*

1.4 Batasan Masalah

Adapun batasan masalah dalam pembuatan perangkat lunak ini adalah sebagai berikut :

1. Koleksi dokumen yang digunakan dalam proses analisis dan implementasi merupakan dokumen berita IT yang telah ditentukan sebelumnya.
2. Dokumen yang digunakan berupa dokumen berita dalam bahasa Indonesia.
3. Perangkat lunak yang dibangun berupa *prototype* aplikasi *Information Retrieval System.*
4. Mengimplementasikan langkah-langkah *Information Retrieval System* pada web.
5. Menghitung kemiripan antara *query* dengan daftar berita yang tersedia.
6. Menganalisis *Information Retrieval System,* meliputi *Stop Word, Tokenisasi, Stemming,* pemberian bobot *term index,* dan model ruang vektor.

7. Model yang digunakan dalam *Information Retrieval System* adalah Model Ruang Vektor.
8. Dokumen Berita berupa teks.

1.5 Metodologi Penelitian

Metodologi Penelitian yang digunakan dalam penulisan skripsi ini adalah sebagai berikut :

a. Studi literatur

Tahap ini dilakukan dengan cara mempelajari literatur-literatur baik yang berupa buku, jurnal, artikel ilmiah, dan *website* yang berhubungan dengan *Information Retrieval System.*

b. Analisis

Pada tahap ini akan dijelaskan proses *Information Retrieval System* dalam hal *Stop Word, Stemming,* pemberian bobot *term index,* dan model ruang vektor.

c. Perancangan

Pada tahap ini akan dilakukan perancangan terhadap perangkat lunak *Information Retrieval System* yang akan dibangun, termasuk menentukan lingkungan pembuatan, bahasa pemrograman yang digunakan, arsitektur, fungsionalitas, dan antarmuka sistem.

d. Implementasi dan Pengujian

Pada tahap ini akan dilakukan implementasi *Information Retrieval System* pada web serta dilakukan pengujian untuk mendapatkan kesimpulan dan saran untuk pengembangan lebih lanjut.

1.6 Sistematika Penulisan

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

BAB I PENDAHULUAN

Menguraikan tentang latar belakang permasalahan, merumuskan inti permasalahan yang dihadapi, menentukan maksud dan tujuan penelitian, yang kemudian diikuti dengan pembatasan masalah, metodologi penelitian serta sistematika penulisan.

BAB II. LANDASAN TEORI

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.

BAB III. ANALISIS DAN PERANCANGAN SISTEM

Membahas tentang analisis sistem, pengguna, serta perancangan sistem untuk pembangunan aplikasi.

BAB IV. IMPLEMENTASI DAN PENGUJIAN SISTEM

Membahas tentang implementasi dan pengujian sistem yang telah dikerjakan serta menerapkan kegiatan implementasi dan pengujian pada perangkat lunak aplikasi.

BAB V. KESIMPULAN DAN SARAN

Berisi kesimpulan dan saran dalam penulisan skripsi.