

# BAB I

## PENDAHULUAN

Dalam bab ini akan dibahas latar belakang dilaksanakannya penelitian, identifikasi masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### 1.1 Latar Belakang

Peranan informasi sangat besar bagi kehidupan dalam berbagai bidang, seperti politik, ekonomi, sosial, budaya, dan lain-lain. Dalam kehidupan masyarakat, teknologi informasi berfungsi untuk memenuhi sebagian besar kebutuhan hidup. Pesatnya perkembangan zaman, kebutuhan informasi di tengah masyarakat mulai berkembang. Kebutuhan masyarakat akan informasi telah menjadikan informasi menjadi kebutuhan yang kontinyu dan rutin, berbagai kebutuhan informasi dimulai dari informasi mengenai berita, surat, cerita, laporan penelitian, data keuangan, dan lain-lain telah menjadi vital bagi masyarakat. Semakin tingginya kebutuhan akan informasi juga akan berdampak pada tingkat peredaran informasi di tengah masyarakat.

Sekitar 3 juta dokumen terindeks di Google, 10-20 *terabytes* teks tersebar di *web*, 1000 *terabytes* informasi (digital & non-digital) dihasilkan setiap tahunnya. Dengan banyaknya informasi yang tersebar luas dapat menyebabkan kualitas pencarian menurun lalu muncul masalah dari cara mengakses informasi menjadi bagaimana memilih informasi yang berguna. Usaha untuk memilih informasi yang memiliki nilai guna inilah yang dinilai masih sulit dilakukan melihat banyaknya informasi yang tersaji.

Beberapa contoh seperti Springer, Amazon, PubMed menyediakan fasilitas untuk mencari informasi dokumen terutama dokumen artikel, literatur, atau jurnal untuk kepentingan studi. Namun tidak semua disediakan secara gratis. Adapun repositori-repositori lokal menyediakan dokumen yang dibutuhkan namun tidak semua didukung fasilitas pencarian yang baik..

Menghadapi akan masalah pencarian dokumen teks yang relevan, dibutuhkan sistem yang dapat mengindeks informasi yang banyak tersaji lalu mengolahnya dalam suatu wadah pengelola informasi. Tujuannya untuk

mempermudah dalam mengelola dan mencari informasi yang relevan dengan sumber terpercaya serta mudah untuk mengaksesnya. Sistem temu kembali informasi merupakan sistem yang dapat diterapkan untuk dapat menemukan informasi yang relevan dari suatu kumpulan informasi secara otomatis (Mandala & Setiawan, 2002).

Dalam pencarian dokumen teks terdapat beberapa metode yang digunakan. Model *Boolean* merupakan model pencarian paling awal dikembangkan dan masih banyak digunakan sampai saat ini. Namun kelemahan model ini menghasilkan pencarian dokumen yang tidak terangkai dan penggunaan *query* pencarian dengan operator matematika. Salah satu metode yang dapat diimplementasikan adalah model ruang vektor. Model ruang vektor dapat diterapkan untuk membantu menyelesaikan permasalahan *retrieval* pada pencarian dokumen teks. Metode ini dapat menghasilkan pencarian yang terangkai serta menggunakan *query* pencarian yang lebih familiar (Salton & McGill, 1983).

Diharapkan dengan dibangunnya sistem temu kembali informasi dengan Implementasi Model Ruang Vektor untuk Pencarian Dokumen Teks, dapat membantu memecahkan masalah pencarian dokumen berbahasa Indonesia. Sehingga diharapkan pengguna dapat dimudahkan dalam mencari dokumen berbahasa Indonesia yang relevan.

## 1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Apakah model ruang vektor yang diterapkan mampu melakukan proses *retrieval* dokumen teks pada sistem?
2. Apakah algoritma *stemming* bahasa Indonesia Nazief Adriani yang diterapkan mampu mengoptimasi proses pengindeksan dokumen pada sistem?
3. Berapa nilai performansi untuk *precision*, *recall*, dan *f-measure* yang diperoleh pada sistem pada saat proses *retrieval*?
4. Apakah metode yang digunakan pada penelitian ini mampu melakukan proses pengolahan dan *retrieval* dokumen teks?

### 1.3 Batasan Masalah

Batasan masalah terhadap penelitian ini adalah sebagai berikut:

1. Data yang dapat diolah adalah dokumen yang berkeekstensi *\*.doc* dan *\*.txt*.
2. Data diproses dari koleksi dokumen yang tersimpan.
3. Sistem yang akan dikembangkan dalam penelitian ini hanya akan berfokus pada proses *retrieval*.
4. Sumber data yang digunakan dalam penelitian ini berasal dari data penulisan karya ilmiah yang berada pada repositori UPI.
5. Data yang diproses merupakan dokumen berbahasa Indonesia.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan model ruang vektor dalam proses *retrieval* pada sistem.
2. Menerapkan algoritma *stemming* bahasa Indonesia Nazief Adriani untuk optimasi proses pengindeksan dokumen pada sistem.
3. Menghitung nilai *precision*, *recall*, dan *f-measure* sistem pada proses *retrieval*.
4. Membuktikan model ruang vektor dan algoritma Nazief Adriani mampu melakukan proses pengolahan dan *retrieval* dokumen teks dengan baik atau tidak.

### 1.5 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan dapat memudahkan pengguna untuk mendapatkan dokumen yang relevan dari sekumpulan dokumen yang tersimpan di koleksi besar berdasarkan masukan kata kunci dari pengguna.

### 1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Dalam menyajikan laporan skripsi ini digunakan sistematika penulisan sebagai berikut:

## BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan.

**BAB II KAJIAN PUSTAKA**

Bab ini menjelaskan dasar teori yang digunakan dalam penelitian skripsi.

**BAB III METODOLOGI PENELITIAN**

Bab ini berisi penjelasan mengenai tahap-tahap yang dilakukan pada proses penelitian.

**BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Bab ini berisi hasil penelitian dan pembahasan penelitian yang dilakukan.

**BAB V KESIMPULAN DAN SARAN**

Bab ini berisi tentang kesimpulan berdasarkan tujuan yang ingin dicapai dalam penelitian skripsi ini dan saran-saran yang dapat diberikan untuk penggunaan dan pengembangan perangkat lunak lebih lanjut.