

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Information Retrieval System adalah sistem pencari yang melakukan proses pencarian informasi secara otomatis berdasarkan isi maupun konteks tertentu sesuai dengan *query* inputan pengguna.

Proses yang dilakukan sistem pencari informasi yang baik, pada umumnya terdiri dari beberapa langkah berikut [2] :

1. *Query pre-processing*, salah satunya adalah proses normalisasi bahasa pada *term*.
2. Proses pembuatan indeks pada pembangunan sebuah sistem pencari.
3. *Post-processing* pada dokumen kandidat yang akan dipilih sebagai dokumen yang dianggap relevan terhadap *query*. Salah satunya dengan meranking dokumen yang relevan terhadap *query*.
4. Perbaikan *query* berdasarkan pada *feedback* dari *user* dan re-evaluasi terhadap *query*.

Langkah tersebut sangat berperan penting terhadap kualitas hasil yang ditampilkan oleh sistem kepada pengguna. Namun secara garis besar sistem pencari informasi terdiri dari dua tahapan yaitu proses *indexing* dan *searching*.

Proses pengindeksan adalah proses mengindeks *term* dari seluruh dokumen yang tersimpan pada sistem. Hasil pengindeksan tersebut disimpan pada basis data indeks dalam bentuk *inverted list* untuk setiap *term*. Dengan kata lain, *inverted list* menyimpan sejumlah *posting(id dokumen)* dimana setiap *term* yang telah terindeks muncul.

Indeks dibuat untuk memudahkan pencarian dokumen. Sehingga sistem tidak perlu membaca satu persatu dokumen untuk menemukan informasi yang diinginkan.

Pada sebuah sistem pencari dengan dokumen yang tersimpan sering berubah (penambahan maupun penghapusan dokumen), maka struktur *inverted list* yang digunakan sebaiknya mendukung *dynamic indexing* (indeks yang bersifat dinamis; adanya penambahan maupun penghapusan *id* dokumen pada *inverted list*).

Pada Tugas Akhir ini dipelajari sebuah struktur data untuk penyimpanan *inverted list* yang memungkinkan *dynamic indexing* yaitu struktur SB-tree.

1.2 Perumusan Masalah

Berdasarkan uraian diatas, maka permasalahan yang muncul dan yang menjadi objek penelitian pada Tugas Akhir ini ialah:

1. Bagaimana mengimplementasikan sebuah struktur penyimpanan *inverted list* yang mendukung *dynamic indexing* untuk sistem pencari dokumen teks dengan dokumen yang sering berubah.
2. Dari struktur yang digunakan tersebut, bagaimana sistem melakukan proses *merge inverted list* untuk menemukan dokumen yang relevan terhadap *query*.

Batasan masalah agar tidak meluasnya materi pembahasan dalam tugas akhir ini ialah:

2. Analisis dilakukan pada *document collection* (tersimpan pada *disk*) yang bersifat statis.
3. *Document collection* pada sistem yaitu berupa *text document*.

1.3 Tujuan

Secara umum tujuan penulisan yang ingin dicapai dalam Tugas Akhir ini ialah:

1. Menerapkan struktur data SB-tree untuk penyimpanan *inverted list* pada sistem pencari dokumen teks yang mendukung dilakukan *dynamic indexing*.
 - menganalisis ukuran basis data yang dibutuhkan untuk menyimpan *inverted list* dengan struktur SB-tree.
 - menganalisis waktu yang dibutuhkan untuk proses update *inverted list* pada struktur SB-tree.
2. Menerapkan algoritma khusus yang memanfaatkan struktur SB-tree untuk melakukan proses *merge inverted list* secara efektif dan efisien.
 - menganalisis waktu yang dibutuhkan untuk *merge inverted list* pada struktur SB-tree.
 - menganalisis pengaruh frekuensi *term* pada *query* terhadap nilai Performance Gain Estimation.
 - menganalisis pengaruh ukuran *block size* sistem terhadap kinerja sistem dalam proses *merge inverted list*.

1.4 Metodologi Penyelesaian Masalah

Metodologi yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini ialah:

1. Studi literatur.
Melakukan diskusi dengan dosen pengajar dan mencari permasalahan yang terjadi pada pemrosesan *indexing* pada sistem pencari teks. Kemudian mencari algoritma sebagai solusinya dan mengumpulkan informasi yang terkait dengan proses *indexing* dan memahami konsep algoritma yang digunakan dari struktur data SB-tree pada proses *indexing* melalui literatur berupa makalah, buku, maupun jurnal berupa dokumen elektronik maupun fisik.

2. Pencarian dan pengumpulan data.
Data yang akan digunakan berupa dokumen elektronik berbahasa Inggris yang diperoleh secara bebas dari internet.
3. Analisis kebutuhan dan perancangan aplikasi yang akan dibangun.
Menganalisa kebutuhan perangkat lunak dan merancang perangkat lunak untuk implementasi struktur data SB-tree pada proses *indexing*.
4. Implementasi dan pengujian
Mengimplementasikan hasil analisis dan perancangan perangkat lunak dengan menggunakan teknik berorientasi objek serta melakukan pengujian dan pengukuran performansi dari sistem.
5. Analisis hasil pengujian dan membuat kesimpulan
Perangkat lunak yang dihasilkan dievaluasi berdasarkan data yang diperoleh dari hasil pengujian.
6. Penyusunan laporan tugas akhir.
Pembuatan laporan tugas akhir yang mendokumentasikan tahap-tahap kegiatan dan hasil dalam tugas akhir ini.

1.5 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini terdiri dari 5 Bab, yaitu:

BAB I Pendahuluan

Bab ini membahas kerangka penelitian dalam tugas akhir, meliputi latar belakang, perumusan masalah, batasan masalah, tujuan perancangan dan metodologi yang digunakan dalam perancangan sistem.

BAB II Landasan Teori

Bab ini menjelaskan seluruh teori yang menjadi landasan konseptual dan mendukung penyelesaian tugas akhir ini.

BAB III Perancangan Sistem

Bab ini membahas mengenai pengumpulan data analisis dan perancangan sistem yang terdiri dari perancangan alur kerja sistem, perancangan basis data, perancangan *interface* dan perancangan modul pada perangkat lunak.

BAB IV Implementasi, Pengujian dan Analisis Sistem

Bab ini membahas implementasi detail sistem, pengujian terhadap sistem dan menganalisis hasil pengujian tersebut untuk mengetahui kesimpulan dari metode yang diimplementasikan.

BAB V Kesimpulan dan Saran

Berisi tentang kesimpulan dan saran yang dapat diambil dari keseluruhan sistem yang telah dibuat.