

Abstraksi

Text Retrieval System adalah sistem pencari dokumen teks berdasarkan *query* masukan pengguna. Sistem pencari dokumen teks terdiri dari dua proses yaitu proses pengindeksan dan pencarian dokumen pada koleksi dokumen di sistem.

Pengindeksan adalah proses mengindeks seluruh *term* dari seluruh dokumen yang tersimpan pada sistem. Pembangunan indeks bertujuan untuk memudahkan sistem menemukan dokumen yang dicari berdasarkan *query*. Karena hasil dari pengindeksan tersebut adalah informasi dokumen (umumnya dikodekan berupa *id*) dimana tiap *term* yang telah terindeks muncul. Sehingga sistem tidak perlu membaca satu persatu dokumen untuk menemukan informasi yang diinginkan (membutuhkan waktu yang banyak dan proses komputasi yang besar [13]). Informasi tersebut tersimpan pada *inverted list*.

Pada sebuah sistem pencari teks dengan dokumen yang tersimpan sering berubah (penambahan maupun penghapusan dokumen), dibutuhkan struktur *inverted list* yang mempercepat proses update *inverted list* untuk mendukung *dynamic indexing*.

Salah satu struktur *inverted list* yang mendukung *dynamic indexing* adalah SB-tree. SB-tree adalah varian spesial dari struktur B(Bayer)-tree yang memiliki informasi tambahan dan format elemen yang spesial pada *leaf node*.

Update *inverted list* dilakukan dengan mengunjungi *node-node* yang mendekati *id* dokumen yang akan diupdate hingga ditemukan *id* dokumen tersebut. Sehingga sistem tidak perlu menelusuri seluruh *id* dokumen secara sekuensial.

Proses pencarian dokumen pada Text Retrieval System dilakukan dengan mencocokkan (*merge*) *inverted list* dari setiap *term* pada *query*. Pada sistem yang mengimplementasikan struktur SB-tree, algoritma yang digunakan untuk proses *merge inverted list*, dimulai dengan mencari irisan dari seluruh root node dari setiap *term* pada *query*. Sistem akan mengunjungi *node* di bawahnya jika *node* tersebut beririsan dengan *node* dari *seluruh* term lainnya, demikian seterusnya hingga ke *leaf node*. Dengan algoritma tersebut kinerja sistem dalam menemukan dokumen menjadi efektif dan efisien.

Kata Kunci: *Text Retrieval, Text Retrieval System, indexing* dan SB-tree.