

Dokumentasi Tugas 2

1. Program java menggunakan lucene yang mengimplementasikan mesin pencari teks sederhana, yang mampu memproses query boolean (AND, OR, dan NOT)

Tahap pertama adalah mempersiapkan file input. File input berisi dokumen dengan masing – masing dokumen terdiri dari tag id, judul, dan teks sebagai content dokumen. Setiap content dokumen diproses menjadi index.

Tahap kedua adalah memproses dokumen ke dalam index, masing – masing dokumen dibagi ke dalam field. Field id, judul, dan teks.

Tahap ketiga adalah proses index (indexing). Kita menambahkan dokumen beserta field ke dalam IndexWriter. IndexWriter berfungsi untuk melakukan indexing, menyimpan indexing ke dalam direktori dan melakukan editing terhadap indexing. Parameter dari IndexWriter adalah analyzer (class : StandardAnalyzer). Analyzer ini yang nantinya akan menganalisa masing – masing dokumen untuk dilakukan indexing sesuai dengan field.

Tahap keempat adalah pencarian (searching). Kita melakukan pencarian query dengan menggunakan index yang sudah dibuat. Untuk melakukan pencarian di index kita menggunakan kelas IndexSearcher dengan parameter IndexReader. IndexReader digunakan untuk membaca hasil indexing. Hasil pencarian kita kumpulkan menggunakan kelas TopScoreDocCollector karena kita hanya menginginkan 10 dokumen pertama yang memiliki bobot tertinggi. Dokumen di diurutkan berdasarkan score masing – masing dokumen.

2. Penggunaan facet menggunakan lucene

Penggunaan facet pada lucene bermaksud untuk melakukan pencarian berdasarkan kategori. Karena kita ingin mencari kelompok penulis dari suatu berita beserta jumlah dokumennya. Maka kita menempatkan penulis sebagai kategori.

Proses Indexing

Untuk menyimpan kategori yang berhubungan dengan index. Kita membuat dua directory berbeda untuk masing masing index dokumen dan kategori. Untuk index dokumen kita menggunakan IndexWriter sama seperti kasus pertama. Untuk menyimpan kategori kita menggunakan DirectoryTaxonomyWriter. Kita membuat kategori penulis di direktoriTaxonomyWriter kemudian menambahkannya ke dalam file index.

Proses Searching

Untuk melakukan pencarian, kita perlu membaca file index dan taxonomy (kategori) yang sudah dibuat. Proses pembacaan dilakukan menggunakan DirectoryReader. Untuk melakukan pencarian, kita menggunakan IndexSearcher dengan parameter DirectoryReader kemudian memanggil method search untuk mencari dokumen yang memenuhi kueri dan menyimpan hasilnya menggunakan kelas FacetsCollector.

Karena yang kita peroleh baru dokumen berdasarkan kueri. Selanjutnya kita perlu menghitung dokumen yang diperoleh berdasarkan kategori penulisnya. Untuk itu kita perlu menghitung berapa jumlah dokumen berdasarkan kategori penulis menggunakan kelasFastTaxonomyFacetCounts dengan parameter FacetCollector dan dile index kategori (taxonomy). Hasil disimpan ke dalam sebuah ArrayList untuk di cetak.