

Soal 03 (Boolean Retrieval Model)

1. Tujuan

Tujuan eksperimen ini adalah membangun sistem pencarian sederhana berbasis Boolean Retrieval Model menggunakan dokumen hasil preprocessing dari sistem rekomendasi buku.

Sistem diharapkan dapat melakukan pencarian menggunakan operator logika AND, OR, dan NOT, serta menampilkan hasil pencarian dokumen yang relevan.

2. Pembangunan Struktur Index

Dari 10 dokumen teks hasil preprocessing, dibentuk:

- **Vocabulary:** kumpulan kata unik dari seluruh dokumen.
Contoh: ['anak', 'pedang', 'hutan', 'cinta', 'motivasi', ...]
- **Inverted Index:** struktur data yang memetakan term ke daftar dokumen tempat term tersebut muncul.
Contoh:

'anak' → ['buku_fantasi.txt']

'pedang' → ['buku_fantasi.txt']

'cinta' → ['buku_romansa.txt']

'motivasi' → ['buku_motivasi.txt']

Struktur index ini digunakan untuk mempercepat pencarian dokumen berdasarkan kata kunci tertentu.

3. Metode Pencarian (Boolean Query Parser)

Sistem menggunakan operator logika Boolean:

- AND → mencari dokumen yang mengandung kedua term.
- OR → mencari dokumen yang mengandung salah satu term.
- NOT → mencari dokumen yang tidak mengandung term tersebut.

Contoh query yang diuji:

1. pedang AND hutan
2. cinta OR motivasi
3. NOT horror

4. Hasil Pengujian

Query	Hasil Sistem	Dokumen Relevan (Gold)	Precision	Recall
pedang AND hutan	buku_fantasi.txt	buku_fantasi.txt	1.00	1.00
cinta OR motivasi	buku_romansa.txt	buku_romansa.txt, buku_motivasi.txt	1.00	0.50
NOT horor	semua kecuali horor	semua kecuali horor	0.90	1.00

Penjelasan :

- Query pedang AND hutan menggunakan operator **AND**, sistem mengambil irisan dua himpunan dokumen yang mengandung kata *pedang* dan *hutan*.
→ hasil: *buku_fantasi.txt*.
- Query cinta OR motivasi mengambil gabungan dua himpunan dokumen yang mengandung *cinta* atau *motivasi*.
→ hasil: *buku_romansa.txt*.
- Query NOT horor mengambil semua dokumen kecuali yang mengandung kata *horor*.
→ hasil: semua file kecuali *buku_horor.txt*.