

# SOAL 05 – Term Weighting, Search Engine, dan Evaluasi

## Tujuan

Menyatukan seluruh modul (preprocessing, indexing, retrieval, dan evaluasi) menjadi sebuah mini search engine yang dapat melakukan pencarian dokumen menggunakan dua model utama:

- Boolean Retrieval
- Vector Space Model (VSM)  
Serta membandingkan dua skema pembobotan istilah yaitu TF-IDF dan BM25 untuk menilai kualitas hasil pencarian.

## Langkah Pengerjaan

### 1. Term Weighting

- Dua skema pembobotan digunakan:
  - **TF-IDF** standar
  - **BM25** (opsional bonus)
- Setiap model diuji menggunakan kumpulan dokumen hasil preprocessing di folder data/processed.

### 2. Search Engine Orchestrator (`search_engine.py`)

- Program dijalankan melalui Command Line Interface (CLI) menggunakan argumen:  
`python src/search_engine.py --model boolean --query "pedang AND hutan"`  
`python src/search_engine.py --model vsm --query "cinta motivasi" --k 3`
- Menampilkan daftar dokumen relevan beserta skor cosine similarity dan kata kunci yang cocok.

### 3. Main Interface (`chat.py`)

- Menyediakan antarmuka interaktif sederhana berbasis teks.
- Pengguna dapat mengetikkan query secara langsung, dan sistem akan menampilkan Top-k dokumen hasil pencarian menggunakan model VSM.

### 4. Evaluasi (`eval.py`)

- Menggunakan Precision, Recall, F1, MAP@k, dan nDCG@k untuk membandingkan performa TF-IDF dan BM25.
- Hasil perbandingan divisualisasikan dalam bentuk grafik batang.

## Hasil Implementasi

### ◆ Search Engine (CLI)

Contoh output:

```
PS D:\TUUUUUGGGGGAAAAASSSSSS\stki-uts-A11.2023.15390-AtanasiusMarcello> python src/search_engine.py --model boolean --
query "pedang AND hutan"

Model: BOOLEAN RETRIEVAL
Query: pedang AND hutan
=====
1. buku_fantasi.txt | anak laki-laki bernama arka menemukan pedang ajaib tersembunyi hutan terlarang bantuan penyihir tua menyelamatkan kerajaan

Total hasil: 1 dokumen.
```

#### ◆ Chat Interface

```
PS D:\TUUUUUGGGGGAAAAASSSSSS\stki-uts-A11.2023.15390-AtanasiusMarcello> python app/chat.py
=====
👤 Mini Search Assistant (VSM-based)
Ketik pertanyaan atau kata kunci Anda (ketik 'exit' untuk keluar)
=====

👤 Query: pedang hutan
👤 Berdasarkan pencarian untuk 'pedang hutan', berikut 3 dokumen teratas:
1. buku_fantasi.txt (cosine: 0.349) - anak laki-laki bernama arka menemukan pedang ajaib tersembunyi hutan terlarang bantuan penyihir tua menyelamatkan kerajaan
2. buku_sains.txt (cosine: 0.000) - penulis fenomena alam mudah dipahami gerhana matahari misteri lubang hitam angkasa
3. buku_romansa.txt (cosine: 0.000) - kisah cinta insan terhalang jarak berjuang mempertahankan hubungan kesibukan ambisi masing-masing

👤 Sistem menampilkan hasil paling relevan berdasarkan kesamaan deskripsi teks.
=====

👤 Query: exit
👋 Terima kasih! Program selesai.
```

#### ◆ Evaluasi Term Weighting

Model	Precision	Recall	F1	MAP@3	nDCG@3
TF-IDF	0.33	0.84	0.47	0.84	0.88
BM25	0.33	0.84	0.47	0.84	0.88

#### ◆ Grafik Perbandingan

