

## **SOAL 01 – Konsep STKI & Perkembangan**

### **1. Definisi STKI dan Perbedaannya dengan Database Retrieval**

Sistem Temu Kembali Informasi (STKI) adalah sistem yang dirancang untuk menemukan dan menyajikan informasi yang relevan dari kumpulan dokumen teks berdasarkan kebutuhan pengguna. Tujuan utama STKI bukan hanya mencari data yang persis sama, tetapi menemukan dokumen yang paling relevan terhadap suatu query atau permintaan informasi.

Berbeda dengan sistem basis data (Database Retrieval), yang bekerja pada data terstruktur seperti tabel, kolom, dan nilai pasti, STKI beroperasi pada data tidak terstruktur seperti teks, artikel, atau deskripsi buku. Sistem basis data menggunakan pencocokan eksak (exact match), sedangkan STKI menggunakan pendekatan aproksimasi (approximate match), yaitu menilai tingkat kesamaan antara query dengan isi dokumen.

Contohnya:

- Database retrieval:  
SELECT \* FROM buku WHERE judul = 'Harry Potter'  
→ hanya menampilkan hasil yang judulnya tepat sama.
- Information retrieval:  
Query: "buku fantasi tentang sekolah sihir"  
→ sistem akan menampilkan buku yang paling relevan, seperti Harry Potter, Percy Jackson, dan sebagainya.

Dua komponen penting dalam STKI adalah:

- Indexing (Inverted Index): struktur data yang menyimpan hubungan antara kata dan dokumen agar pencarian lebih cepat.
- Ranking: proses mengurutkan dokumen berdasarkan tingkat relevansinya terhadap query, biasanya menggunakan metode pembobotan seperti TF-IDF dan pengukuran kemiripan seperti Cosine Similarity.

### **2. Garis Besar Arsitektur Search Engine Klasik**

Secara umum, arsitektur search engine klasik terdiri atas beberapa tahap utama:

1. Korpus Dokumen  
Sumber data berupa kumpulan teks atau deskripsi dokumen (misalnya deskripsi buku).
2. Preprocessing  
Melakukan pembersihan teks agar siap diolah: case folding, tokenisasi, stopword removal, dan stemming.
3. Indexing (Inverted Index)  
Membangun struktur data yang menyimpan kata-kata penting dan dokumen tempat kata itu muncul.

#### 4. Query Processing

Input pengguna juga diproses dengan langkah yang sama seperti dokumen agar bentuknya seragam.

#### 5. Retrieval & Matching

Sistem mencocokkan query dengan index menggunakan model retrieval (misalnya Boolean atau VSM).

#### 6. Ranking

Dokumen diurutkan berdasarkan nilai kesamaan (relevance score).

#### 7. Presentation

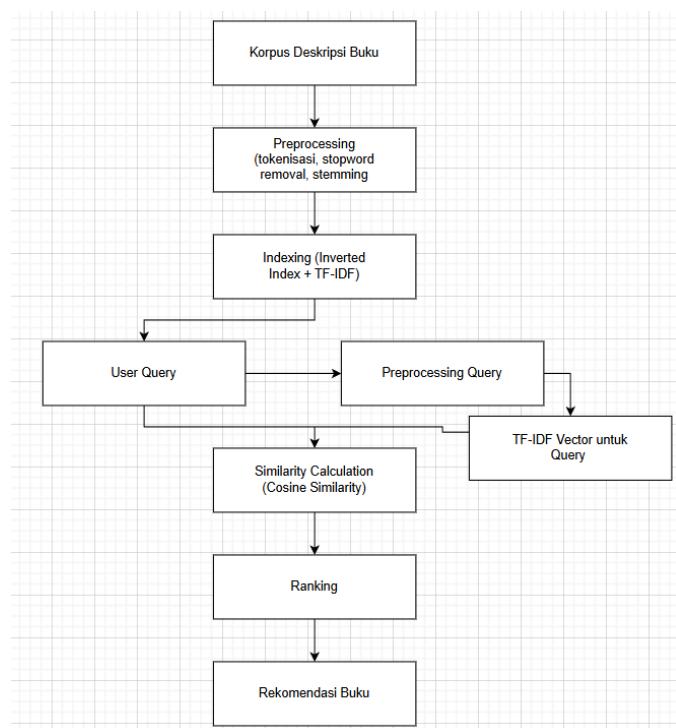
Menampilkan hasil pencarian yang paling relevan kepada pengguna.

Alur umumnya:

Query → Preprocess → Retrieve → Rank → Present

### 3. Sketsa Arsitektur Retrieval Klasik (Vector Space Model)

Diagram arsitektur sistem rekomendasi buku berbasis teks ditunjukkan sebagai berikut:



Pada arsitektur ini, terdapat dua jalur paralel:

- Jalur kiri untuk korpus dokumen, yang diolah dan diindeks satu kali di awal.
- Jalur kanan untuk query pengguna, yang diproses setiap kali pencarian dilakukan.

Keduanya kemudian bertemu pada tahap Similarity Calculation, di mana sistem menghitung tingkat kesamaan antara deskripsi buku dan query menggunakan Cosine Similarity.

#### 4. Kaitan Soal 02 – 05 dengan Materi RPS

Soal	Kaitan Materi STKI (RPS)
Soal 02 – Document Preprocessing	Tahap pembersihan teks (case folding, tokenisasi, stopword removal, stemming).
Soal 03 – Indexing dan Inverted Index	Membangun struktur index untuk pencarian cepat.
Soal 04 – Term Weighting & Vectorization	Menghitung TF-IDF dan membentuk vektor representasi dokumen dan query.
Soal 05 – Retrieval & Evaluation	Menggunakan Cosine Similarity untuk menghitung kesamaan dan evaluasi hasil.