

Nama : Aditya Rendy Setyawan

NIM : A11.2023.15189

Mata Kuliah : Sistem Temu Kembali Informasi

Kelas : A11.4703

SOAL 01 - Konsep STKI & Perkembangan (15%, Sub-CPMK10.1.1)

1. Definisi STKI

Sistem Temu Kembali Informasi (STKI) adalah sistem yang dirancang untuk menemukan informasi relevan dari sekumpulan dokumen berdasarkan *kebutuhan informasi pengguna (user query)*.

Berbeda dengan *database retrieval* yang menekankan pada pencocokan tepat (exact match) terhadap data terstruktur menggunakan bahasa kueri seperti SQL, STKI bekerja dengan data tidak terstruktur (teks, artikel, dokumen), menggunakan pendekatan pencarian relevansi (relevance-based matching).

Aspek	Database Retrieval	Sistem Temu Kembali Informasi
Data	Terstruktur (tabel)	Tidak terstruktur (teks)
Pencarian	Berdasarkan kesamaan tepat (exact match)	Berdasarkan kemiripan makna (similarity)
Bahasa Query	SQL	Natural language / keyword
Hasil	Akurat tapi terbatas	Relevan tapi bisa bervariasi

Dua komponen utama STKI adalah:

- Indexing: proses representasi dokumen menjadi bentuk vektor atau daftar term agar mudah dicari.
- Ranking: penentuan urutan dokumen berdasarkan tingkat kemiripan dengan query pengguna.

2. Garis Besar Arsitektur Search Engine Klasik

Arsitektur STKI modern umumnya terdiri dari lima tahap utama:

1. Document Processing (Preprocessing)
Dokumen dikumpulkan dan diproses (case folding, tokenisasi, stopword removal, stemming).
Hasilnya berupa kumpulan token siap diindeks.
2. Indexing
Sistem membuat inverted index berisi daftar term dan dokumen yang memuatnya.
Ini mempercepat pencarian dibanding membaca seluruh teks mentah.
3. Query Processing
Query dari pengguna diproses dengan cara yang sama seperti dokumen (tokenisasi dan stemming).

4. Retrieval & Ranking

Sistem melakukan pencarian dengan model tertentu:

- *Boolean Model*: menggunakan operator AND, OR, NOT.
- *Vector Space Model (VSM)*: menghitung kemiripan (cosine similarity) antara query dan dokumen.

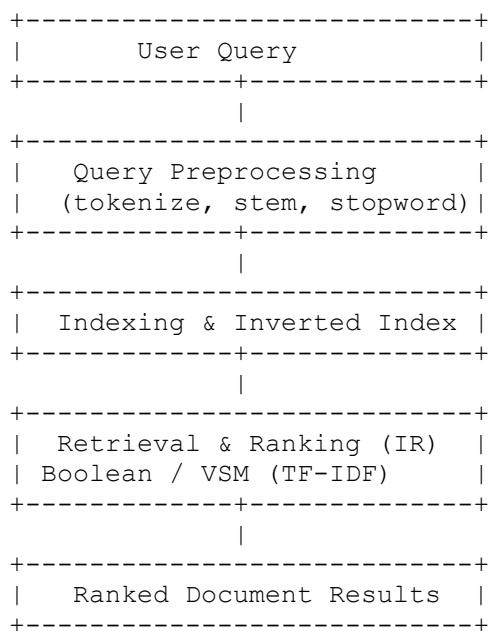
5. Presentation

Hasil pencarian ditampilkan berurutan berdasarkan skor relevansi tertinggi.

Alur umum:

Query → Preprocessing → Matching (Boolean/VSM) → Ranking → Presentasi

3. Sketsa Arsitektur Retrieval Klasik



4. Peta Materi ke RPS STKI

Soal	Topik Materi STKI	Sub-CPMK
Soal 02	<i>Text Preprocessing (tokenisasi, stopword, stemming)</i>	10.1.2
Soal 03	<i>Boolean Retrieval Model & Inverted Index</i>	10.1.3
Soal 04	<i>Vector Space Model (TF-IDF, Cosine Similarity)</i>	10.1.4
Soal 05	<i>Evaluasi IR (Precision, Recall, MAP, nDCG)</i>	10.1.5
Soal 01	<i>Konsep dasar & arsitektur sistem IR klasik</i>	10.1.1