

Soal 01 - Konsep STKI & Perkembangan

Sistem Temu Kembali Informasi (STKI) merupakan cabang ilmu komputer yang berfokus pada pencarian dan pengambilan informasi relevan dari sekumpulan data atau dokumen berdasarkan permintaan (query) pengguna (Vika Ummu Hani, 2025). Fokus utama STKI adalah *similarity matching*, yaitu menemukan dokumen yang paling mirip atau paling relevan dengan kebutuhan pengguna, bukan sekadar pencocokan literal. Search Engine seperti Google, Bing, dan e-Library Udinus adalah contoh paling umum penerapan STKI. Adapun perbedaan Information Retrieval dengan Database Retrieval yakni sebagai berikut :

Aspek	Information Retrieval	Database Retrieval
Tujuan Utama	Menemukan dokumen paling relevan berdasarkan kemiripan informasi.	Mengambil data yang <i>exact match</i> sesuai kondisi query.
Jenis Data	Data <i>unstructured</i> atau <i>semi-structured</i> (teks, artikel, berita).	Data <i>structured</i> (tabel, kolom, baris).
Basis Pencocokan	<i>Similarity matching</i> (TF-IDF, cosine similarity, BM25).	<i>Exact matching</i> (SQL: =, >, <, LIKE).
Kemampuan Menangani Bahasa Alami	Tinggi karena dapat menangani stemming, sinonim, konteks, tokenisasi.	Rendah karena tidak memahami makna bahasa, hanya cocok literal.
Hasil Pencarian	Berupa daftar dokumen yang di- <i>ranking</i> berdasarkan relevansi.	Berupa record yang cocok secara presisi tanpa ranking.

Setelah memahami perbedaan IR dan Database Retrieval, dua komponen kunci dalam STKI adalah index dan ranking. Index berfungsi mempercepat pencarian dengan menyimpan daftar dokumen yang memuat suatu term, sehingga sistem tidak perlu membaca seluruh dokumen. Sementara itu, ranking digunakan untuk mengurutkan dokumen berdasarkan relevansinya terhadap query menggunakan bobot seperti TF-IDF, cosine similarity, atau BM25. Dengan ranking, hasil pencarian tidak sekadar “cocok/tidak cocok”, tetapi tersusun dari yang paling relevan sehingga lebih informatif bagi pengguna. Adapun arsitektur dasar sistem STKI terdiri dari tahap:

1. Corpus Dokumen: Sekumpulan teks yang akan dicari (missal: publikasi dari Udinus)
2. Document Preprocessing: Mengubah teks mentah menjadi format melalui case folding, tokenisasi, stopwords removal, dan stemming.

3. Indexing: Menemukan dokumen terkait dengan cepat tanpa membaca seluruh isi dokumen.
4. Query Processing: Input keyword atau query, lalu diproses seperti dokumen.
5. Retrieval: Sistem mencocokkan query dengan dokumen dan menghasilkan daftar dokumen yang mengandung atau berkaitan dengan query.
6. Ranking: Jika menggunakan VSM atau model berbobot lainnya, sistem menghitung tingkat relevansi setiap dokumen terhadap query.
7. Presentation: Menampilkan hasil pencarian kepada pengguna dalam bentuk daftar.

Adapun sketsa arsitektur retrieval klasik (Boolean dan VSM) sebagai berikut:



Adapun peta materi ke RPS yakni :

Soal	Materi	Sub-CPMK	Penjelasan
02	Preprocessing Dokumen (tokenisasi, stemming, dan penghapusan stopword)	10.1.2	Menangani proses awal pengolahan teks sebagai bagian dari pipeline IR.
03	Model Boolean Retrieval	10.1.3	Menerapkan pencarian berbasis operasi logika seperti AND, OR, dan NOT.
04	Vector Space Model (VSM) dan TF-IDF	10.1.3	Melakukan penilaian relevansi dokumen dan menghasilkan urutan (ranking) berdasarkan bobot istilah.
05	Evaluasi Sistem (precision, recall, F1-score, dan metrik lainnya)	10.1.4	Menilai kualitas kinerja model pencarian menggunakan metrik evaluasi IR.

