

Soal 01

Sistem Temu Kembali Informasi (STKI) adalah sistem yang dirancang untuk mencari dan mengambil informasi yang relevan dari kumpulan dokumen yang tidak terstruktur, seperti teks, artikel, atau berita, untuk memenuhi kebutuhan informasi pengguna. STKI berfokus pada proses pencarian informasi, bukan pencarian data; yang dicari bukan hanya data mentah, tetapi makna dan isi informasi yang sesuai dengan kebutuhan pengguna. STKI mencakup dua proses utama dalam konteks teknis: indexing (pembuatan indeks dari dokumen teks) dan retrieval (pencarian keterangan atau dokumen berdasarkan query pengguna). Search engine seperti Google, Bing, dan e-Library Udinus adalah contoh paling umum dari penerapan STKI, yang dapat mencari dokumen teks dari kumpulan data berukuran besar dengan cepat dan efisien.

Aspek	Database Retrieval	Sistem Temu Kembali Informasi
Data yang dicari	Terstruktur (tabel, field)	Tidak terstruktur (dokumen, teks)
Bahasa query	Formal (SQL)	Natural (kata kunci biasa)
Hasil pencarian	Harus cocok 100% dengan query	Boleh mengandung ketidaktepatan, tetapi relevan
Kesalahan kecil	Dapat menyebabkan kegagalan query	Masih bisa ditoleransi
Contoh	<i>SELECT * FROM mahasiswa WHERE nim='A11.2023.15263'</i>	“Cari artikel tentang kecerdasan buatan”

Tujuan utama STKI adalah menemukan dokumen yang relevan sebanyak mungkin, dan meminimalkan dokumen yang tidak relevan. Karena relevansi bersifat subjektif, sistem STKI biasanya menggunakan ukuran seperti term frequency, inverse document frequency, dan cosine similarity untuk menentukan tingkat kemiripan antara dokumen dan query.

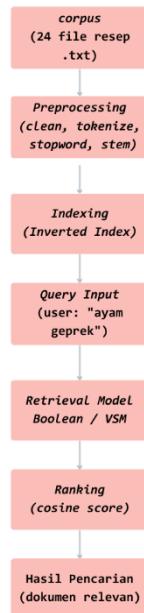
Dalam STKI, index berperan seperti “daftar isi cepat” untuk mempercepat proses pencarian. Proses indexing mengubah teks menjadi struktur data seperti inverted index, yang menyimpan setiap kata dan dokumen yang mengandung kata tersebut.

Sementara itu, ranking menentukan urutan hasil pencarian berdasarkan relevance score, misalnya dengan metode Vector Space Model (VSM) atau BM25. Tujuannya agar dokumen paling relevan muncul di posisi teratas hasil pencarian.

Arsitektur dasar sistem STKI terdiri dari beberapa tahap berikut:

1. Korpus Dokumen: sekumpulan teks yang akan dicari (misalnya kumpulan resep masakan).

2. Preprocessing: pembersihan teks melalui case folding, tokenizing, stopword removal, stemming, dan normalisasi tanda baca.
 3. Indexing: membangun inverted index agar pencarian lebih cepat.
 4. Query Processing: pengguna mengetikkan kata kunci, lalu query diproses seperti dokumen.
 5. Retrieval & Ranking: sistem mencari dokumen relevan menggunakan model seperti Boolean atau Vector Space Model (VSM), lalu menghitung skor kemiripan.
 6. Presentation: menampilkan hasil pencarian dengan urutan relevansi.
- Sketsa Arsitektur Retrieval Klasik (Boolean dan VSM)



- Peta Materi ke RPS (Soal 02 – 05)

Mapping Soal terhadap Sub-CPMK			
Soal	Materi	Sub-CPMK	Keterangan
Soal 02	Document Preprocessing (tokenisasi, stemming, stopword)	10.1.2	Implementasi tahap awal pipeline IR
Soal 03	Boolean Retrieval Model	10.1.3	Penerapan pencarian berbasis logika AND/OR/NOT
Soal 04	Vector Space Model (VSM) dan TF-IDF	10.1.3	Mengukur relevansi dan membuat ranking dokumen
Soal 05	Evaluasi dan Analisis Hasil (precision, recall, F1-score)	10.1.4	Mengukur performa sistem pencarian

