

Nama : Suluh Yoga Pratama
NIM : A11.2023.14979
Kelompok : A11.4703

Konsep STKI & Perkembangan

I. Definisi STKI dan Perbedaannya dengan Database Retrieval

Sistem Temu Kembali Informasi (STKI) adalah sistem yang dirancang untuk menemukan dan menampilkan dokumen yang paling relevan dengan kebutuhan informasi pengguna (query) dari sekumpulan dokumen yang tidak terstruktur. Berbeda dengan database retrieval yang bekerja pada data terstruktur (seperti tabel dan kolom dalam SQL) dan mencari pencocokan eksak, STKI bekerja pada data tidak terstruktur seperti teks dan bertujuan menemukan relevansi makna.

Sebagai contoh, pada database retrieval, query seperti `SELECT * FROM mahasiswa WHERE nama='Budi'` hanya akan menampilkan baris yang tepat sama. Sedangkan dalam STKI, jika pengguna mengetik query “pencarian informasi mahasiswa”, sistem akan menampilkan dokumen yang mengandung makna serupa misalnya teks tentang data mahasiswa, sistem akademik, atau profil kampus. STKI juga memiliki dua komponen penting:

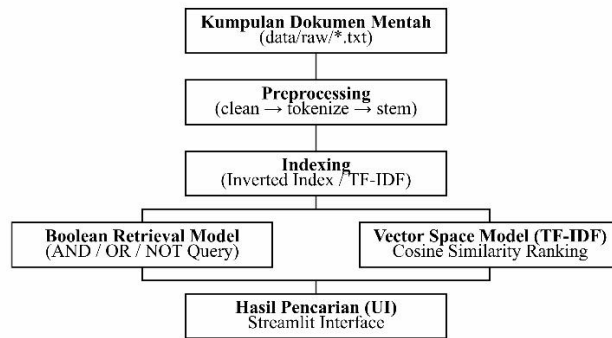
1. Index: struktur data yang mempercepat pencarian kata kunci dalam sekumpulan dokumen, seperti inverted index.
2. Ranking: proses pemberian skor relevansi untuk menentukan urutan hasil (misalnya menggunakan TF-IDF dan cosine similarity).

II. Arsitektur Klasik Sistem Pencarian Informasi

Arsitektur sistem STKI secara umum terdiri dari beberapa tahapan utama yang menggambarkan alur kerja dari korpus dokumen hingga hasil pencarian ditampilkan ke pengguna. Berikut urutan prosesnya:

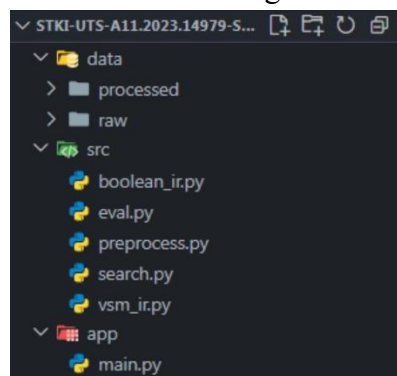
1. Document Collection (Korpus): kumpulan dokumen teks mentah (pada proyek ini disimpan di data/raw/*.txt).
2. Preprocessing: pembersihan dan normalisasi teks meliputi case folding, tokenization, stopword removal, dan stemming untuk menghasilkan token yang seragam (data/processed/*.txt).
3. Indexing: sistem membangun struktur inverted index dan menghitung bobot TF-IDF agar pencarian bisa dilakukan cepat.
4. Query Processing: pengguna mengetik query yang kemudian diproses sama seperti dokumen agar formatnya seragam.
5. Retrieval Model: sistem mencocokkan query dengan dokumen menggunakan model pencarian:
6. Boolean Model: pencarian berbasis logika AND, OR, NOT.
7. Vector Space Model (VSM): pencarian berbasis bobot TF-IDF dan perhitungan cosine similarity.
8. Ranking: dokumen diurutkan dari yang paling relevan ke paling tidak relevan.
9. Presentation: hasil pencarian ditampilkan ke pengguna (melalui antarmuka Streamlit).

III. Sketsa Arsitektur STKI (Retrieval Klasik Boolean & VSM)



Gambar 1. Sketsa Arsitektur STKI

Gambar ini menggambarkan alur sistem STKI klasik mulai dari korpus teks → preprocessing → indexing → retrieval → ranking → hasil pencarian.



Gambar 2. Struktur Folder Proyek STKI

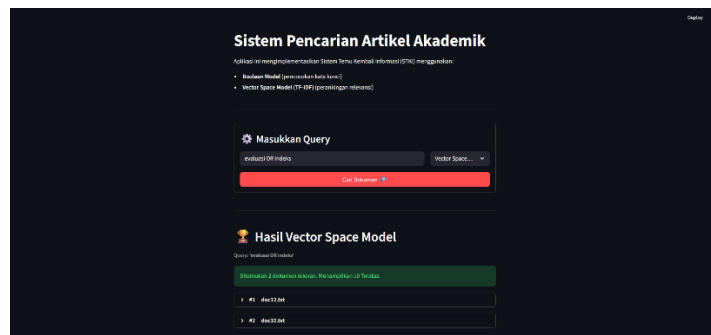
Gambar ini menunjukkan bagaimana sistem dibangun modular sesuai arsitektur STKI klasik (data → proses → retrieval → UI).

```
D:\@Yoga\Muliah\Sem_5\SistemTentuKembaliInformasi\stki-uts-A11.2023.14979-SuluhYogaPratama>python -c "from src.preprocess import process_folder; process_folder('data/raw','data/processed')"
```

```
[OK] doc1.txt → processed (64 tokens)
[OK] doc10.txt → processed (63 tokens)
[OK] doc11.txt → processed (63 tokens)
[OK] doc12.txt → processed (63 tokens)
[OK] doc13.txt → processed (63 tokens)
```

Gambar 3. Hasil Proses Preprocessing di Terminal

Tampilkan log seperti [OK] doc1.txt → processed (64 tokens). Gambar ini menunjukkan tahap awal arsitektur, yaitu konversi korpus mentah menjadi token siap index.



Gambar 4. Tampilan Streamlit Saat Query Dijalankan

Masukkan query, lalu tampilkan hasil pencarian (baik Boolean maupun VSM). Gambar ini menunjukkan bagian akhir arsitektur, yaitu presentation layer tempat hasil ranking ditampilkan ke pengguna.

IV. Peta Materi ke RPS (Keterkaitan Soal 02–05)

Tabel 1. Peta Materi ke RPS

Soal	Topik/Tujuan	Keterkaitan Materi RPS
Soal 2	Document Preprocessing	Materi 2 – Document Preprocessing
Soal 3	Boolean Retrieval Model	Materi 3 – Model STKI Dengan Boolean Model
Soal 4	Vector Space Model & Ranking	Materi 4 – Model STKI Vector Space Model
Soal 5	Evaluasi & Integrasi Search Engine	Materi 7 – Evaluasi & Search Engine Concept

V. Kesimpulan

Sistem Temu Kembali Informasi (STKI) adalah fondasi utama dari teknologi pencarian modern seperti Google, di mana prosesnya dimulai dari pengumpulan dokumen hingga menampilkan hasil yang paling relevan. Pada proyek ini, STKI diimplementasikan menggunakan dua model klasik Boolean dan Vector Space dengan pipeline lengkap dari preprocessing, indexing, retrieval, ranking, hingga evaluasi. Implementasi ini membantu memahami bagaimana setiap komponen saling terhubung dan membentuk sistem pencarian yang efisien dan terukur.