NAMA : Nur Kholis

NIM : A11.2022.14584

SOAL 01 – KONSEP STKI & PERKEMBANGANNYA

Bobot: 15% | Sub-CPMK 10.1.1

1. Definisi STKI dan Perbedaannya dengan Database Retrieval

Sistem Temu Kembali Informasi (STKI) adalah bidang ilmu komputer yang berfokus pada bagaimana sistem dapat menemukan dokumen yang relevan terhadap kebutuhan pengguna dari kumpulan data yang besar, khususnya pada data tidak terstruktur seperti teks. Berbeda dengan sistem basis data (*database retrieval*) yang menggunakan pendekatan pencarian eksak (*exact match*) melalui query SQL, STKI bekerja dengan konsep relevansi (*best match*). Dalam basis data, hasil pencarian bersifat deterministik—suatu data hanya akan muncul jika memenuhi seluruh kondisi query—sedangkan dalam STKI hasil pencarian bersifat probabilistik dan diurutkan berdasarkan nilai kesamaan atau skor relevansi. Dua komponen kunci dalam STKI adalah index dan ranking function. Struktur *inverted index* berfungsi memetakan istilah ke daftar dokumen yang mengandung istilah tersebut sehingga mempercepat pencarian, sedangkan fungsi peringkat (*ranking function*) seperti *Term Frequency–Inverse Document Frequency (TF-IDF)* dan *cosine similarity* digunakan untuk menilai dan mengurutkan dokumen sesuai tingkat kesesuaiannya dengan query pengguna. Dengan kombinasi kedua komponen tersebut, STKI tidak hanya menampilkan dokumen yang sesuai, tetapi juga mengoptimalkan urutan hasil berdasarkan relevansi yang paling tinggi.

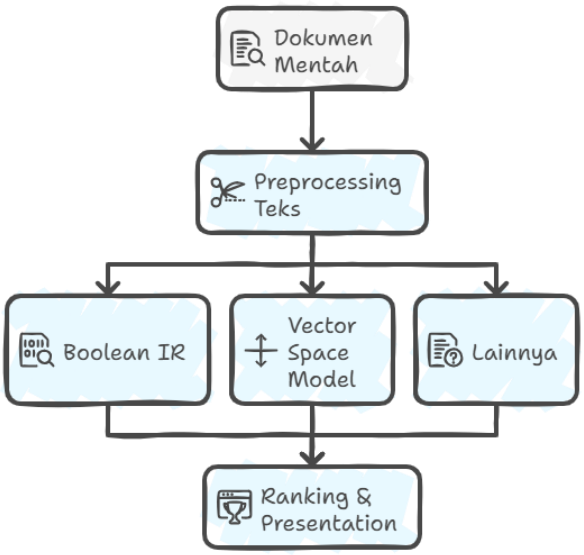
2. Garis Besar Arsitektur Search Engine Klasik

Arsitektur *search engine* klasik menggambarkan alur sistem temu kembali informasi yang bekerja secara sistematis dari tahap pengumpulan data hingga penyajian hasil kepada pengguna. Tahapan pertama adalah document collection, yaitu pengumpulan dokumen dari berbagai sumber data seperti artikel, web, atau arsip digital. Selanjutnya dilakukan preprocessing, yang bertujuan menormalkan teks melalui proses *case-folding*, tokenisasi, penghapusan *stopword*, dan *stemming* agar kata memiliki bentuk dasar yang seragam. Setelah itu, sistem memasuki tahap indexing, di mana dibangun struktur *inverted index* atau matriks bobot TF-IDF yang digunakan untuk mempercepat pencarian. Pada tahap query processing, input pengguna diproses agar sesuai dengan representasi dokumen yang telah diindeks. Kemudian sistem melakukan retrieval dan

ranking, yaitu proses pencarian dan pengurutan dokumen berdasarkan skor relevansi menggunakan model seperti Boolean atau *Vector Space Model (VSM)*. Akhirnya, tahap presentation menampilkan hasil pencarian berupa daftar dokumen (*result list*) yang telah diurutkan dari yang paling relevan hingga paling rendah. Dengan arsitektur ini, sistem STKI mampu mengelola proses temu kembali informasi secara efisien, fleksibel, dan relevan terhadap kebutuhan pengguna.

3. Sketsa Arsitektur Retrieval Klasik (Boolean vs VSM)

Model *Boolean Retrieval* dan *Vector Space Model (VSM)* merupakan dua pendekatan utama dalam sistem temu kembali informasi klasik. Model Boolean bekerja dengan logika himpunan dan menggunakan operator seperti AND, OR, serta NOT untuk menentukan apakah suatu dokumen memenuhi ekspresi query. Model ini efisien untuk pencarian eksak, namun tidak dapat memberikan peringkat berdasarkan tingkat relevansi. Sebaliknya, model *Vector Space Model* memperlakukan setiap dokumen dan query sebagai vektor dalam ruang multidimensi, di mana setiap dimensi merepresentasikan istilah unik dalam koleksi. Bobot setiap term dihitung menggunakan metode TF-IDF, dan kesamaan antara query serta dokumen dihitung menggunakan *cosine similarity*. Semakin besar nilai cosine, semakin relevan dokumen terhadap query. Gambaran umum arsitektur retrieval klasik dapat dijelaskan sebagai berikut:



Gambar 1. Skema arsitektur

4. Peta Materi ke RPS dan Hubungan dengan Soal 02–05

Soal	Topik Ujian	Materi RPS / Sub-CPMK Terkait	Fokus Kompetensi
01	Konsep STKI & Arsitektur	Materi 1–2 (Konsep, Index, Ranking)	Pemahaman konsep & kerangka kerja
02	Preprocessing Dokumen	Materi 3 (Preprocessing teks: tokenisasi, stopword, stemming)	Implementasi text cleaning
03	Boolean Retrieval	Materi 4 (Representasi index dan query)	Logika AND/OR/NOT
04	Vector Space Model	Materi 5–6 (TF-IDF, Cosine Similarity)	Representasi vektor & ranking
05	Evaluasi Sistem IR	Materi 7 (Precision, Recall, MAP, nDCG)	Analisis kinerja & pembobotan