



FAKULTAS  
**ILMU  
KOMPUTER**

CSCE604135 • Temu-Balik Informasi  
Semester Genap 2024/2025  
Fakultas Ilmu Komputer, Universitas Indonesia

**Tugas Kelompok Turu, Brok? Iya**

Rhaken Shaquille Akbar Yanuanda - 2206814791

Reyhan Wiyasa Puspanegara - 2206081925

Hanan Adipratama - 2206081824

**Ketentuan**

1. Untuk Tugas Kelompok ini, Anda dapat memilih salah satu dari dua pilihan tugas.
2. Tugas Kelompok dikerjakan secara berkelompok, dimana setiap kelompok terdiri dari dua hingga tiga orang. Daftar kelompok Anda di [sini](#).
3. Terdapat dua deadline untuk tugas ini, yakni deadline progres dan deadline final.
4. Kumpulkan tugas pada submisi yang telah disediakan di SCeLE sebelum:
  - Deadline Progres: Jumat, 9 Mei 2025 23:59 WIB
  - Deadline Final: [to be announced]
5. Perhatikan kelengkapan untuk submisi sesuai panduan pengumpulan. Pengumpulan yang tidak lengkap atau tidak sesuai akan beresiko untuk tidak dinilai.
6. Tugas ini dirancang sebagai tugas yang dikerjakan per kelompok. Plagiarisme antar kelompok tidak diperkenankan dalam bentuk apapun. Adapun kolaborasi berupa diskusi (tanpa menyalin maupun mengambil jawaban orang lain) dan literasi masih diperbolehkan dengan mencantumkan kolaborator dan sumber.
7. Anda boleh konsultasi dengan tim asisten dosen. Asisten dosen diperbolehkan membantu Anda dengan memberikan petunjuk.

# Progress Report

## Pilihan 2: Implementasi Sistem Search Engine

Repo: <https://github.com/hanan-collab/TK-TuruBro-Iya/tree/main>

Hari, Tanggal	Versi	Deskripsi
Jum'at, 9 Mei 2025	0.0.1	<p><b>Progress</b></p> <p><b>a. Arsitektur dan Implementasi Sistem</b> Sistem telah berhasil dipecah menjadi dua komponen utama:</p> <ul style="list-style-type: none"><li>- Jupyter Notebook (tk_tbi.ipynb) yang bertanggung jawab atas proses indexing embedding teks pertanyaan dari dataset ke dalam Elasticsearch.</li><li>- Backend API menggunakan FastAPI (backend.py) yang menyediakan endpoint /search/ untuk melakukan pencarian semantik berbasis teks dari pengguna.</li></ul> <p>Docker Compose telah disiapkan untuk menjalankan layanan Elasticsearch secara lokal, mempermudah pengembangan dan pengujian.</p> <p><b>b. Integrasi Model Machine Learning</b> Model all-MiniLM-L6-v2 dari sentence-transformers digunakan untuk menghasilkan embedding vektor dari pertanyaan.</p> <p>Embedding ini digunakan baik untuk proses indexing maupun pencarian berbasis cosine similarity di Elasticsearch.</p> <p><b>c. Pengolahan dan Penyimpanan Data</b> Dataset diambil dari <code>ir_datasets.load("beir/quora")</code> yang kemudian di ubah ke dataframe dan dikonversi ke .csv.</p> <p>Dataset quora_dataset.csv digunakan untuk melatih dan mengisi indeks Elasticsearch dengan representasi vektor dari pertanyaan.</p> <p>Proses indexing dilakukan melalui notebook dengan menyimpan teks dan embedding ke dalam field Elasticsearch (text, doc_id, embedding).</p> <p><b>d. Dokumentasi</b> Progress di dokumentasikan di repository berikut, <a href="https://github.com/hanan-collab/TK-TuruBro-Iya/tree/main">https://github.com/hanan-collab/TK-TuruBro-Iya/tree/main</a></p>

		<p><b>Plan</b></p> <p><b>a. Tampilan FrontEnd</b> Merancang UI sederhana agar pengguna dapat melakukan pencarian tanpa menggunakan API langsung.</p> <p><b>b. Deployment</b> Menyusun strategi deployment menggunakan layanan cloud seperti Railway, Vercel (untuk frontend), atau server VPS untuk Elasticsearch dan backend.</p>
--	--	--

**Terima kasih!**