



FAKULTAS  
**ILMU  
KOMPUTER**

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

### **Tugas Kelompok**

**Tenggat Waktu Progres: Jumat, 9 Mei 2025 23:59 WIB**

**Tenggat Waktu Final: [to be announced]**

### **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 SCell 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.

## Petunjuk Pengerjaan Tugas

Terdapat dua pilihan tugas yang dapat dipilih oleh setiap kelompok, yakni dalam bentuk riset atau pengembangan sistem search engine yang akan dijelaskan di lebih lanjut di sectionnya masing-masing. Secara umum, berikut ini kriteria yang perlu dipenuhi untuk tugas kelompok:

- Anda **tidak diperbolehkan menggunakan codebase dari tugas pemrograman sebelumnya**.

### Pilihan 1: Riset IR

Pada opsi ini, Anda diharapkan bisa melakukan penelitian sederhana terkait bidang IR. Secara umum, terdapat dua pendekatan yang dapat dipilih:

1. Anda dapat merancang dan mengembangkan pendekatan atau model baru dalam IR, ataupun melakukan modifikasi terhadap model yang sudah ada.
2. Anda dapat mereproduksi paper penelitian dalam bidang IR, terutama yang sudah dipublikasikan di konferensi atau jurnal terkemuka, seperti SIGIR, ACL, ECIR, WSDM, dan sebagainya. Untuk mengakses paper berbayar, manfaatkan akses langganan perpustakaan UI melalui <https://remote-lib.ui.ac.id/>.

Output dari pilihan ini berupa sebuah paper singkat 4 halaman yang berisi laporan riset dan kode eksperimen Anda. Pastikan kodenya **rapi dan readable (beserta dokumentasi yang lengkap)**, serta eksperimennya **reproducible** (menggunakan versi library dan random seeds yang fixed). Jika Anda menggunakan Jupyter Notebook, gunakan markdown cells untuk menambahkan penjelasan agar lebih terstruktur. Sementara itu, paper Anda diharapkan mencakup hal-hal berikut.

Bagian	Penjelasan
Judul dan Penulis	Judul yang relevan dan nama anggota kelompok.
Pendahuluan	Latar belakang dan masalah yang ingin diselesaikan.
Metodologi	Penjelasan model, arsitektur, atau pendekatan yang digunakan.
Eksperimen	Dataset yang digunakan, metrik evaluasi, dan konfigurasi eksperimen.
Hasil dan Analisis	Hasil eksperimen dan analisisnya.
Kesimpulan	Ringkasan temuan dan potensi pengembangan selanjutnya.
Referensi	Referensi yang digunakan dalam pelaksanaan eksperimen.

## Pilihan 2: Implementasi Sistem Search Engine

Pada opsi ini, Anda diharapkan membangun sistem praktis berbasis search engine seperti yang telah dipelajari dan diimplementasikan pada tugas pemrograman sebelumnya. Terdapat beberapa ketentuan pada opsi ini (selain ketentuan umum), yakni sebagai berikut.

- Anda wajib melakukan eksplorasi terkait teknologi search engine terkini seperti [Elasticsearch](#) atau [OpenSearch](#) untuk diimplementasikan ke dalam proyek Anda. Dilarang menggunakan codebase tugas pemrograman sebelumnya.
- Anda wajib mengembangkan sistem search engine end-to-end dengan tampilan, misalnya dalam bentuk website seperti Google Search. Silakan berikan nama (*brand*) untuk sistem search engine Anda.
- Anda **wajib menggunakan LLM** dalam tugas ini, untuk kebutuhan apapun itu. Anda bebas memilih LLM yang akan digunakan, selama LLM tersebut bersifat **open** (contohnya adalah LLMs yang weights-nya di-publish di HuggingFace). Terdapat beberapa opsi yang dapat Anda pilih, baik itu untuk LLM yang di-host secara local maupun yang bukan. Khusus untuk yang local, dikarenakan kebutuhan compute power yang besar, silakan gunakan LLM dengan jumlah parameter yang lebih kecil ( $\leq 0.5B$ ), seperti [Qwen2.5-0.5B-Instruct](#). Berikut adalah opsi-opsinya.
  - Anda dapat menggunakan beberapa LLM gratis yang dapat diakses menggunakan API melalui HuggingFace, seperti [Meta-Llama-3-8B-Instruct](#), [Mistral-7B-Instruct-v0.3](#), dan sebagainya. Perlu diperhatikan bahwa penggunaan API untuk free tier memiliki rate limit per harinya, sehingga gunakanlah dengan bijak. Sebagai referensi, silakan membaca halaman [ini](#) terkait Serverless Inference API HuggingFace. Selain melalui HuggingFace, Anda juga dapat menggunakan API jika disediakan secara gratis oleh pengembang LLM tersebut.
  - Anda dapat menggunakan REST API dari model-model yang disediakan oleh [Ollama](#). Perhatikan bahwa Ollama ini juga bersifat local.
  - Anda dapat membuat API Anda sendiri menggunakan [FastAPI](#) untuk serve pemanggilan LLM local.
  - Opsi lain yang memungkinkan untuk Anda.
- Anda bisa menentukan sendiri koleksi data yang ingin Anda gunakan. Koleksi data boleh dalam bahasa Inggris maupun bahasa Indonesia, dan boleh berasal dari domain lain (misalkan domain medis). Anda boleh menggabungkan beberapa koleksi data menjadi satu. Berikut contoh dataset:
  - NFCorpus (domain medis): [Link](#)
  - Berbagai *dataset* yang tertera pada [ir-datasets.com](#)

Silakan lakukan eksplorasi sumber dataset lain seperti HuggingFace, Kaggle, maupun yang lainnya. Sertakan sumbernya di dokumentasi kode Anda.

Output dari pilihan ini berupa video YouTube dan kode sumber. Video YouTube memuat setidaknya berisi penjelasan arsitektur sistem, teknologi yang digunakan, dan demo singkat penggunaan aplikasi. Kode diunggah ke suatu repository di platform version control seperti GitHub atau Gitlab secara publik. Sertakan sebuah file README untuk dokumentasi, yang

setidaknya berisi penjelasan sistem dan arsitektur, cara instalasi dan menjalankan sistem, teknologi dan dataset yang digunakan, serta contoh penggunaan aplikasi.

## Pengumpulan Progres

Untuk semua pilihan tugas kelompok, kumpulkan sebuah *file* .zip dengan format **TKP\_NPM1\_NPM2\_[NPM3].zip** (contoh: **TKP\_2106130001\_2006130002.zip**) yang berisi:

- File .pdf yang berisi laporan singkat sekitar 1-2 halaman terkait apa yang sudah dikerjakan dan apa yang direncanakan untuk dikerjakan kedepannya.
- File .py, .ipynb, atau kumpulan file source code pekerjaan Anda (dimasukkan ke dalam suatu folder).

## Pengumpulan Final

Kumpulkan sebuah *file* .zip dengan format **TKF\_NPM1\_NPM2\_[NPM3].zip** (contoh: **TKF\_2106130001\_2006130002.zip**) yang berisi:

- Pilihan 1:
  - File .py atau .ipynb kode eksperimen.
  - File .pdf laporan eksperimen.
- Pilihan 2:
  - File .txt yang berisi link ke video YouTube penjelasan aplikasi (pastikan dapat diakses, bebas apakah ingin dibuat public atau unlisted).
  - File .txt yang berisi link ke online repository yang berisi kode sistem (pastikan visibility repository di-set public).

**Selamat mengerjakan!**