

# INSTITUT TEKNOLOGI NASIONAL

Jalan P.H.H Mustafa 23 Bandung 40124

## Evaluasi Akhir Semester Ganjil 2025/2026

Mata Ujian	: IFB-307 Data Mining & IR	Tanggal	: 12 Januari 2026
Jurusan	: Informatika	Waktu	:
Dosen	: FA	Sifat	: Take home test

### Proyek Akhir IFB-307 – Information Retrieval System

#### 1. Deskripsi Singkat

Pada Proyek Akhir IFB-307, mahasiswa diminta membangun sebuah **Sistem Information Retrieval (IR)** berbasis teks untuk topik tertentu. Sistem harus mampu menerima query, menampilkan hasil pencarian yang relevan, serta menyediakan ringkasan dokumen.

Proyek dikerjakan secara **berkelompok (2 mahasiswa per kelompok)**.

#### 2. Capaian Pembelajaran yang Dievaluasi

Proyek ini mengevaluasi:

- **Sub-CPMK 4:** Konsep IR, preprocessing, inverted index, dan IR model
- **Sub-CPMK 6:** Seleksi fitur teks dan text document summarization

#### 3. Ketentuan Umum

- Topik bebas tetapi **spesifik dan konsisten**
- Dataset minimal **±30 dokumen teks**
- Sumber data harus jelas (dataset publik / scraping etis)
- Bahasa pemrograman bebas (Python dianjurkan)

#### 4. Fitur Wajib Sistem

Sistem harus memiliki:

1. **Text preprocessing** (tokenization, stopword removal, stemming/lemmatization)
2. **Inverted index**
3. **Information Retrieval model** (misalnya TF-IDF + cosine similarity)
4. **Query processing & document ranking**
5. **Seleksi fitur teks**
6. **Text document summarization (extractive)**

#### 5. Deliverable

##### A. Sistem

- Dapat didemonstrasikan
- Menerima query dan menampilkan:
  - Ranking dokumen
  - Ringkasan dokumen

##### B. Dokumentasi Proyek

Format dokumentasi menggunakan format dokumentasi standard profesional pengembangan sistem (professional system development documenation) – lihat lampiran

#### 6. Demonstrasi

- Setiap kelompok wajib:
  - Menjelaskan konsep sistem
  - Mendemonstrasikan sistem secara langsung

#### 7. Catatan Penting

- Fokus penilaian pada **pemahaman konsep dan implementasi**, bukan UI
- Library boleh digunakan, **namun konsep harus dipahami**
- Plagiarisme **tidak ditoleransi**