

# PROJEK AKHIR MATA KULIAH

## INF-501 PENELUSURAN INFORMASI

### GANJIL 2023/2024

#### Dosen Pengasuh

Prof. Dr. Taufik Fuadi Abidin, S.Si., M.Tech

Dr. Nizamuddin, M.Sc

Razief Perucha Fauzie Afidh, S.Si., M.Sc

[taufik.abidin@usk.ac.id](mailto:taufik.abidin@usk.ac.id)

---

## PENDAHULUAN

Projek mata kuliah **Penelusuran Informasi** berkaitan dengan implementasi inverted indeks menggunakan bahasa pemrograman C, melakukan perbandingan hasil dengan pembangkitan inverted indeks, dan merancang prototipe sistem penelusuran informasi berbasis web, menggunakan engine yang diimplementasikan menggunakan bahasa pemrograman C, **swish-e**, atau **Nutch**. Projek akhir mata kuliah ini **berbobot 30%** dari nilai akhir dan sekaligus sebagai **Ujian Akhir Semester (UAS)** mata kuliah ini. Setelah mengerjakan Projek ini, mahasiswa diharapkan dapat:

1. Memahami cara kerja sistem penelusuran informasi, diantaranya cara membangun indeks (inverted indeks), stopword removal, dan melakukan pemeringkatan (*ranking*).
2. Memahami cara menggunakan **engine swish-e** atau **nutch** sebagai back-end.
3. Memahami cara implementasi inverted indeks dengan bahasa C dan mengasah kemampuan melakukan *debugging* terhadap kode program orang lain.
4. Mengetahui cara menghubungkan sisi *front-end* berbasis web dengan sisi back-end (*c-code*, *swish-e*, atau *nutch*).
5. Menghitung waktu yang dibutuhkan untuk membangun indeks dan menentukan peringkat dokumen.
6. Membuat prototipe sebuah sistem pencari.

Projek ini merupakan **kegiatan individu**. Mahasiswa tidak dibenarkan memberikan hasil pekerjaannya kepada mahasiswa yang lain. Semua sumber kode program dari orang lain yang digunakan dalam projek ini harus dirujuk dalam **Laporan Projek**.

## PERMASALAHAN

Ada beberapa hal yang harus dilakukan dalam menyelesaikan projek ini, yaitu:

1. Pelajari dan pahami implementasi proses **pembangkitan inverted indeks** dan **ranking** yang dirancang menggunakan bahasa pemrograman C (code dapat diunduh melalui laman sistem e-learning). Pemahaman code dapat dimulai dengan memahami dan mempelajari file ***index-tools.c*** dan ***query-with-doclen.c***
2. Ubah file *stoplist* agar berisi daftar *stopword* dalam Bahasa Indonesia.
3. *Compile code* menggunakan target yang telah didefinisikan dalam file ***Makefile***.  
Contoh: *make index* atau *make query*
4. Bangkitkan inverted indeks dengan menjalankan file binary *index-db* dan lakukan ranking dengan memasukkan *query* (kata kunci) melalui *command line* menggunakan file binary *querydb*. Gunakan dokumen yang ada dalam direktori **data** dan bangun inverted indeks menggunakan dokumen-dokumen dalam direktori tersebut.

5. Rancang dan implementasikan sebuah halaman sisi depan (*front-end*) ala mesin pencari. Melalui halaman *front-end* tersebut, pengguna dapat memasukkan query sebagai input dalam bentuk *keyword*. Halaman *front-end* ini mirip seperti halaman depan Google, Yahoo, Bing, atau Ask dan memiliki sebuah *text-field* untuk menginputkan query berupa keyword, pop-list untuk menentukan *top k* hasil ranking (nilai *k* adalah 10, 20, 30, 40, dan 50), pilihan *engine* yang akan digunakan (implementasi *c-code*, *swish-e*, atau *nutch*), dan sebuah tombol Cari (*Search*). Ada boleh mengimplementasikan *front-end* menggunakan PHP, CGI, JSP, Perl, Python, atau bahasa script/pemrograman yang lain.
6. Setelah *query* diinputkan, halaman *front-end* akan meneruskan *query* dan nilai *k* ke *engine* disisi *back-end* yang dipilih (default adalah *back-end* implementasi *c-code*). Selanjutnya pemeringkatan dilakukan dan hasilnya pemeringkatan *top k* dikirim ke halaman *front-end* dan ditampilkan. Setiap hasil mohon dapat disertakan tautan (*link*) ke dokumen asli sehingga pengguna dapat mengklik tautan tersebut untuk melihat dokumen aslinya, sama seperti tautan pada hasil mesin pencari di klik.
7. Bandingkan hasil ranking menggunakan engine hasil implementasi menggunakan *c-code*, *swish-e*, atau *nutch*. Analisa hasil perbandingan ditulis secara rinci dalam laporan.
8. Buat video dari presentasi anda terkait Projek ini. Gunakan *software camtasia* atau sejenisnya. Durasi video **maksimum 10 menit** dan unggah video ke Youtube. Tulis dan sertakan URL dari video tersebut dalam **Laporan Projek** sehingga dapat dengan mudah diakses dan diberi penilaian.

**File laporan projek dalam format PDF diunggah melalui sistem e-learning [elearning.usk.ac.id](http://elearning.usk.ac.id).** Letakkan data (di luar direktori *public\_html*) dan prototipe hasil implementasi projek ini server yang mungkin dapat diakses secara publik (jika ada). Sertakan dalam laporan URL dari prototipe mesin pencari yang diimplementasikan.

## WAKTU PENGUMPULAN

Projek ini dikumpulkan paling lambat pada tanggal **16 Desember 2023, jam 23.59 WIB** secara online via sistem e-learning. Dokumen yang harus dikumpulkan adalah softcopy dari **Laporan Projek** dalam format PDF yang ditulis pada kertas ukuran A4, font 12 Times New Romans, margin 1" semua sisi, dan spasi *single* antar baris.



***yes we can!***  
(c) 2023