

Pembuatan Mesin Pencari pada Korpus Teks Bahasa Indonesia

Deskripsi Project: Project ini dikerjakan secara berkelompok yang terdiri dari 2-3 mahasiswa per kelompok. Setiap kelompok diminta untuk membangun sebuah mesin pencari sederhana untuk topik tertentu (*vertical search engine*) untuk korpus teks Bahasa Indonesia. Project ini diselesaikan dalam waktu 2-3 Minggu dan harus dipresentasikan pada pertemuan ke 15 dan 16. Pada pertemuan ke 10 setiap kelompok harus menyiapkan proposal untuk topik yang ingin digunakan pada proyek akhir (ikuti template [laporan TA](#) untuk isi proposal).

Langkah-langkah yang harus dilakukan dalam project ini adalah sebagai berikut:

1. **Pengumpulan Korpus:**

Kumpulkan korpus teks bahasa Indonesia minimal 300 dokumen dengan topik yang sama melalui proses crawling. Jenis dokumen dapat berupa artikel berita, blog, jurnal, skripsi, dan dokumen lainnya yang setiap dokumen mengandung beberapa paragraf.

2. **Pengindeksan:**

- Lakukan pengindeksan terhadap Korpus. Ini melibatkan proses pembersihan teks, tokenisasi (pemisahan teks menjadi kata-kata atau token), penghapusan stopwords, dan stemming (penghapusan awalan dan akhiran kata sehingga menjadi kata dasar). Anda dapat menggunakan algoritma Sastrawi atau algoritma stemming bahasa Indonesia lainnya.
- Indeks semua dokumen ke dalam struktur data pilihan Anda yang memungkinkan pencarian paling efisien.

3. **Pencarian Informasi:** Implementasikan 2 algoritma pencarian informasi yang berbeda, seperti Jaccard *Similarity*, *Vector Space Model*-TFIDF, algoritma BM25, Sentence-BERT, dan lain-lain. Gunakan algoritma ini untuk melakukan pencarian dalam korpus terindeks berdasarkan kueri pencarian yang diberikan.

4. **Evaluasi:** Lakukan perhitungan evaluasi kinerja algoritma yang telah dipilih dengan menggunakan metrik evaluasi yang sesuai, seperti *Precision*, *Recall*, *F1 Score*, *Mean Average Precision* (MAP), dan lainnya. Buat perbandingan hasil evaluasi antara satu algoritma dengan yang lainnya dan jelaskan melalui laporan tertulis.

5. **Pembuatan UI mesin pencari:** Buat tampilan yang menarik untuk *user* interface mesin pencari tersebut dengan tools yang dikuasai untuk menunjukkan perbandingan performa kedua algoritma.
6. **Presentasi Project:** Project wajib dipresentasikan oleh seluruh anggota tim melalui video yang diupload melalui youtube, kemudian live demo pada pertemuan ke 15 dan 16. Semua tahapan yang dikerjakan harus dijelaskan dengan baik. Setiap Langkah yang ditempuh akan dihargai dengan bobot nilai yang sesuai. Setiap kelompok diberikan waktu hanya 5 menit untuk live demo dan sesi tanya jawab.
7. **Laporan Tugas:** Buat laporan tugas yang mencakup tahapan pengerjaan, hasil evaluasi, *screenshot* hasil kerja, kesimpulan, dan referensi. Pastikan untuk menjelaskan setiap tahapan yang dikerjakan dan alasan pemilihan setiap algoritma karena ini menjadi poin penting dalam penilaian. Laporan dikumpulkan paling telat pada pertemuan ke 16 melalui e-learning oleh ketua kelompok saja dengan format Kelompok_angka.pdf (contoh: Kelompok_3.pdf).
8. **Kerja Sama Tim:** Project ini mewajibkan semua anggota kelompok aktif berpartisipasi dalam penyelesaian project dan presentasi hasil. Bagi anggota tim yang tidak bekerja, maka **namanya tidak boleh dicantumkan** di dalam laporan. Pada halaman terakhir laporan harus disebutkan **kontribusi dari setiap anggota kelompok**.

Penilaian:

- Keberhasilan proses crawling dan pembuatan vertical search engine
- Kualitas proposal & laporan
- Estetika bahan tayang (slideshow maupun GUI)
- Penyampaian pada presentasi

Presentasi dinilai secara individu, maka seluruh anggota harus turut andil dalam pemaparan dan demo project.

---Selamat Bekerja---