**Kelompok 4 (1A-D IV)**

* Atika Khoirunnisa (171524006)
* Farhan Rahmat Abdillah (171524010)
* Sophia Gianina Daeli (171524029)

**Deskripsi Aplikasi**

Aplikasi yang akan dibuat adalah aplikasi pemeriksa kemiripan dua buah teks atau lebih (maksimal 5 file teks) dengan berdasarkan perbandingan banyaknya kata yang sama pada teks-teks yang diujikan. File yang dapat diujikan adalah file berbahasa Inggris dengan ekstensi .txt. Aplikasi akan dibuat dalam Bahasa C. Tujuan mengetahui persentase kemiripan teks. Penjelasan spesifikasi program (mis., kategori sililarity ?)

Proses yang akan dilakukan untuk membuat program ini, terdiri atas tahapan-tahapan berikut :

1. Case Folding.

Proses yang dilakukan untuk menghilangkan tanda baca dan menyelaraskan seluruh karakter ke dalam *lower case*.

1. Pembuatan Stopword Tree.

Proses untuk membuat *tree* yang elemennya adalah anggota dari *stopword.* Daftar program stopword diambil dari internet.

1. Stemming.

Proses mengubah setiap kata berimbuhan menjadi kata dasar. ADT untuk *stemming* diambil dari internet.

1. Tokenizing.

Proses memisahkan kata demi kata dalam *file* dan memasukkannya ke dalam *tree*. *Tree* yang dibuat adalah *tree* AVL.

1. Stopword Removal.

Proses ini dilakukan bersamaan dengan *tokenizing.*

1. Menghitung Statistik.

Proses ini dilakukan bersamaan dengan *tokenizing.*

1. Menghitung Persentase Kemiripan.

Proses menentukan presentase dari total frekuensi kemunculan kata yang sama dari teks yang diujikan dengan teks pembanding.

Alur Proses nya ?

**Rancangan Kerja**

|  |  |
| --- | --- |
| **Minggu ke-…** | **Target / Kegiatan** |
| 1 | * Analisis *program requirement* (struktur data, modul-modul, dan ADT). * Memulai pembuatan program. |
| 2 | Pembuatan program. |
| 3 | Uji coba dan evaluasi program |

**Pembagian Tugas**

|  |  |
| --- | --- |
| **Nama** | **Tugas** |
| Atika Khoirunnisa | * Menghitung presentase kemiripan. * Memahami ADT s*temming.* |
| Farhan Rahmat A. | * Melakukan *tokenizing* dengan mengabaikan *stopword* (*stopword removing*) * Memahami ADT s*temming.* |
| Sophia Gianina D. | * *Case folding.* * Membangung *stopword tree*. * Memahami ADT s*temming.* |