**Laporan Deskripsi Program *Words Similarity Check***

PROGRAM STUDI DIPLOMA IV TEKNIK INFORMATIKA

Di Jurusan Teknik Komputer dan Informatika



Oleh

Cahya Ramadhan (171524007)

Imanda Syahrul Ramadhan (171524013)

Muhammad Reyhan Soeharto (171524023)

POLITEKNIK NEGERI BANDUNG

Jl. Gegerkalong Hilir Ds. Ciwaruga, Bandung 40012, Telp: (022)2013789 Fax(022)2013889

**Deskripsi dan Cara Kerja Program**

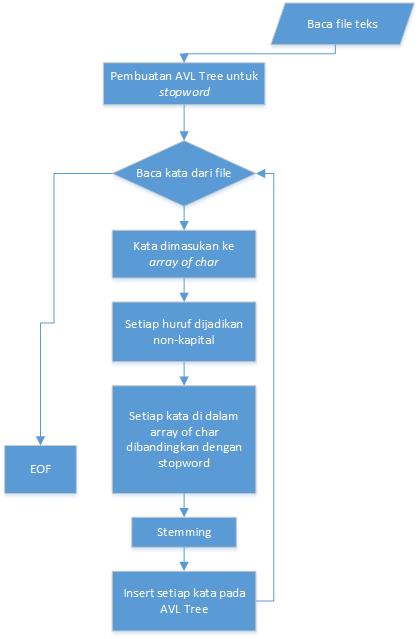
**Deskripsi Program**

Program *Words Similarity Check* ini membantu untuk memeriksa persentasi kesamaan kata dalam suatu *file* terhadap *file* lainnya. Target dari pengguna program ini adalah mahasiswa, guru, dosen, dan sebagainya. Diharapkan dengan dibuatnya program ini menjadikan penggunanya dapat meminimalisir kesamaan kata dalam tulisan yang dikerjakan.

Program akan menanyakan dua buah *file* dengan *file* pertama sebagai bahan uji dan *file* kedua sebagai bahan penguji. Persentasi kesamaan kata dihasilkan dari berapa jumlah kata-kata yang sama dari *file* pertama terhadap kata-kata dari *file* kedua dibagi dengan total kata dari *file* kedua dikalikan dengan seratus.

(Deskripsi program mengenai spesifikasi program (mis. mampu membandingkan berapa file teks bahasa…, definisi similarity, fitur…))

**Cara Kerja Program**



Gambar 1 Flowchart prorese pembacaan kata-kata dari file teks

1. Program akan meminta nama *file* pertama, lalu kata-kata pada *file* pertama akan masuk proses *preprocessing*, dengan cara sebagai berikut. (Gambar 1 menunjukkan proses ini).
   1. Pembuatan AVL Tree yang berisikan kata *stopword*.
   2. Pembacaan kata dari *file*, jika karakter alfabet, dimasukkan ke dalam array of char dan huruf dari kata-kata tersebut dijadikan non-kapital.
   3. Setiap kata yang sudah dimasukkan ke dalam *array of char* dibandingkan dengan setiap kata *stopword*, jika kata berupa *stopword* abaikan dan kembali pada tahap b.
   4. Proses *stemming* dilakukan pada kata yang lolos pada proses perbandingan dengan *stopword*. Proses *stemming* adalah proses penghapusan kata yang berimbuhan menjadi kata dasar.
   5. Proses pembuatan *AVL Tree* untuk setiap kata yang sudah melalui proses – proses di atas, jika kata belum ada pada *tree*, *insert* kata tersebut ke dalam *tree*.
2. Program akan meminta nama *file* kedua, lalu kata-kata pada *file* kedua akan melalui proses *preprocessing* yang sama dengan setiap kata pada *file* pertama.
3. Setelah terdapat *AVL Tree* yang berisi kata-kata beserta dengan frekuensi kemunculannya dari masing-masing *file,* dilakukanlah proses traversal pada *tree* untuk dilakukan pemeriksaan frekuensi pada setiap kata yang sama, lalu diambil frekuensi terkecil dan dimasukkan ke dalam variabel integer dan sambil menghitung total kata.
4. Persentase kesamaan kata dihasilkan setelah mendapatkan jumlah frekuensi dan total kata, kedua variabel tersebut dijadikan bahan perhitungan persentase dengan cara membagi jumlah frekuensi dengan total kata dikali seratus.

Ketika membahas alur proses, hal yang harus diperhatikan adalah input – proses – output setiap tahapannya.

**Timeline**

Tabel 1 menunjukkan rencana dalam pengerjaan Program Text Similiarity.

Table 1 Rencana Timeline dalam mengerjakan program text similiarity

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Kegiatan | | Waktu Pengerjaan | | |
| Minggu I | Minggu II | Minggu III |
| 1 | Pre-Processing | Pencarian kamus stopword |  |  |  |
| 2 | Pencarian kamus kata dasar |  |  |  |
| 3 | Pencarian dan penyesuaian program stemming |  |  |  |
| 4 | Processing | Implementasi tokenizing |  |  |  |
| 5 | Implementasi stopword removal |  |  |  |
| 6 | Implementasi pembuatan statistik kata |  |  |  |
| 7 | Implementasi penghitungan kemiripan file |  |  |  |
| 8 | Post-Processing | Uji Coba Program |  |  |  |
| 9 | Debugging Program |  |  |  |
| 10 | Pembuatan Laporan Akhir |  |  |  |

**Pembagian Tugas**

Tabel 2 menunjukkan rencana pembagian tugas masing-masing anggota kelompok untuk menyelesaikan Program Text Similiarity.

Tabel 2 Rencana pembagian tugas anggota kelompok

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | | |
| **Cahya** | **Imanda** | **Reyhan** |
| 1 | Pencarian kamus kata dasar | Pencarian dan penyesuaian program stemming | Pencarian kamus stopword |
| 2 | Implementasi tokenizing | Implementasi stopword removal | Implementasi pembuatan statistik kata |
| 3 | Implementasi penghitungan kemiripan file | Implementasi penghitungan kemiripan file | Implementasi penghitungan kemiripan file |
| 4 | Uji Coba Program | Uji Coba Program | Uji Coba Program |
|  |  |  |  |
| 5 | Debugging Program | Debugging Program | Debugging Program |
| 6 | Pembuatan Laporan Akhir | Pembuatan Laporan Akhir | Pembuatan Laporan Akhir |

**Struktur Data dan Spesifikasi Modul**

**Struktu Data**

Struktur-struktur data yang akan dipakai untuk memenuhi kerja program adalah sebagai berikut.

infotype : struct [

words[MAKS] : char

freq[2] : integer ]

dengan MAKS = 50

node : struct [

info : infotype

height : integer

right, left : ↑of node ]