# bab iV

# hasil dan pembahasan

## 4.1. Lingkungan Percobaan

Agar sistem yang telah dikembangkan dapat berjalan dengan semestinya, dibutuhkanlah perangkat dengan spesifikasi tertentu, adapun dalam penelitian ini menggunakan spesifikasi perangkat diantaranya:

### 4.1.1. Spesifikasi Perangkat Keras

Perangkat keras yang mendukung aplikasi ini berjalan dengan baik sebagai berikut:

1. *Processor* : Intel(R) Pentium(R) CPU 2020M @ 2.40GHz
2. *RAM* : 6 GB DDR3
3. *Harddisk* : 500 GB
4. *VGA* : AMD HD8570 2GB dedicated

### 4.1.2. Spesifikasi Perangkat Lunak

Perangkat lunak yang mendukung aplikasi ini berjalan dengan baik sebagai berikut:

1. Sistem Operasi : Windows 7 Ultimate 64-bit
2. IDE : Visual Studio Code v.1.38.1
3. DBMS : Mysql Database
4. Web Server : Apache (XAMPP v3.2.2, PHP versi 7.3.0)
5. Browser : Google Chrome

## 4.2. Implementasi Metode dan Langkah Pengujian

### 4.2.1. Data Dokumen

Data dokumen merupakan tahap untuk pengumpulan *dataset* dan proses pengujian data tes. Data yang digunakan berupa dokumen abstrak skripsi mahasiswa Budi Luhur dalam bentuk *pdf* yang kemudian di *convert* dalam bentuk teks. Data dokumen didapatkan dari perpustakaan Universitas Budi Luhur. Data dokumen yang sudah berbentuk teks akan dilakukan tahap *preprocessing* pembentukan *n-gram, rolling hash,* pembentukan *window* dan pencarian *fingerprint.* Sehingga diketahui tingkat *similarity* dari sebuah dokumen.

### 4.2.2. Tahapan *Preprocessing*

Setelah data dokumen didapat, maka dokumen akan memasuki tahap *preprocessing* pada dokumen yang sudah menjadi teks, proses yang dilakukan yaitu, *casefolding,* menghapus karakter selain a-z, mangganti *slangword,* menghapus *stopword.* Berikut bisa dilihat di Tabel 4.1 berikut.

Tabel 1 4.1 *Tahapan Preprocessing*

|  |  |  |
| --- | --- | --- |
| **Keterangan** | **Data Awal** | **Hasil *Preprocessing*** |
| *Dataset* | IMPLEMENTASI AUGMENTED REALITY MENGGUNAKAN METODE MARKER BASED TRACKING BERBASIS ANDROID SEBAGAI MEDIA PROMOSI SEKOLAH PADA SMK YADIKA 4 KARANG TENGAH | implementasiaugmentedrealitymetodemarkerbasedtrakingberbasisandroidpromosismkyadikakarangtengah |
| Data Tes | IMPLEMENTASI TEKNOLOGI AUGMENTED REALITY SEBAGAI MEDIA PEMBELAJARAN DENGAN METODE MARKER BASED TRACKING BERBASIS MOBILE ANDROID PADA PAUD ANANDA KREO SELATAN | implementasiteknologiaugmentedrealitypembelajaranmetodemarkerbasedtrackingberbasismobileandroidpaudanandakreoselatan |

### 4.2.3. Pembentukan *N-Gram*

Setelah dokumen melalui tahap *preprocessing,* kemudian masuk ke tahap pembentukan *n-gram* yaitu, proses memecahkan *string* teks yang dikelompokan berdasarkan nilai *n-gram* yang akan dihitung pergeserannya secara terus menurus ke depan sejumlah nilai n sampai akhir dokumen. Contoh tahapan pembentukan kgram dengan nilai N=3. Berikut bisa dilihat di Tabel 4.2 berikut.

Tabel 2 4.2 Pembentukan *N-Gram*

|  |  |  |
| --- | --- | --- |
| **Keterangan** | **Hasil *Preprocessing*** | ***N-Gram*** |
| *Dataset* | implementasiaugmentedrealitymetodemarkerbasedtrakingberbasisandroidpromosismkyadikakarangtengah | imp mpl ple lem eme men ent nta tas asi sia iau aug ugm gme men ent nte ted edr dre rea eai ali lit ity tym yme met eto tod ode dem ema mar ark rke ker erb rba bas ase sed edt dtr tra rak aki kin ing ngb gbe ber erb rba bas asi sia isa san and ndr dro roi oid idp dpr pro rom omo mos osi sis ism smk mky kya yad adi dik ika kak aka kar ara ran ang ngt gte ten eng nga gah |
| Data Tes | implementasiteknologiaugmentedrealitypembelajaranmetodemarkerbasedtrackingberbasismobileandroidpaudanandakreoselatan | imp mpl ple lem eme men ent nta tas asi sit ite tek ekn kno noi oio iog ogi gia iau aug ugm gme men ent nte ted edr dre rea eal ali lit ity typ ype pem emb mbe bel ela laj aja jar ara ran anm nme met eto tod ode dem ema mar ark rke ker erb rba bas ase sed edt dtr tra rac ack cki kin ing ngb gbe ber erb rba bas asi sis ism smo mob obi bil ile lea ean and ndr dro roi oid idp dpa pau aud uda dan ana nan and nda dak akr kre reo eos ose sel ela lat ata tan |

### 4.2.4. Algoritme *Winnowing*

Algoritme *winnowing* merupakan salah satu algoritme yang berfungsi sebagai dokumen *fingerprint* atau algoritme yang digunakan untuk mendeteksi tindakan plagiarisme dengan menggunakan teknik *hashing*. *Input* dari algoritme *winnowing* berupa dokumen teks, dan akan menghasilkan keluaran berupa kumpulan nilai *hash* yang terbentuk dari peritungan ASCII pada setiap karakter. Berikut tahapanya.

### Perhitungan *Rolling Hash*

Setelah pembentukan *n-gram* maka proses selanjutnya yaitu, perhitungan *rolling hash* untuk mencari nilai *hash* dari setiap *string* yang sudah dipotong pada tahap *n-gram*.Setiap *string* yang sudah dipotong diubah menjadi ASCII lalu dihitung dengan rumus *hash.* Misal kita mengambil dari potongan kalimat *dataset* dan data tes yang ada. Rumus *hash:*

…(4.1)

c1 = nilai ascii dari huruf pertama dari satu buah kgram

b = bilangan prima

k = nilai *k-gram*

Tabel 3 4.3 Kode ASCII

|  |  |  |  |
| --- | --- | --- | --- |
| **Karakter** | **Nilai Unicode** | **Nilai ANSI ASCII** | **Keterangan** |
| (heksadesimal) | (desimal) |
| a | 61 | 97 | Huruf latin b kecil |
| b | 62 | 98 | Huruf latin b kecil |
| c | 63 | 99 | Huruf latin c kecil |
| d | 64 | 100 | Huruf latin d kecil |
| e | 65 | 101 | Huruf latin e kecil |
| f | 66 | 102 | Huruf latin f kecil |
| g | 67 | 103 | Huruf latin g kecil |
| h | 68 | 104 | Huruf latin h kecil |
| i | 69 | 105 | Huruf latin i kecil |
| j | 006A | 106 | Huruf latin j kecil |
| k | 006B | 107 | Huruf latin k kecil |
| l | 006C | 108 | Huruf latin l kecil |
| m | 006D | 109 | Huruf latin m kecil |
| n | 006E | 110 | Huruf latin n kecil |
| o | 006F | 111 | Huruf latin o kecil |
| p | 70 | 112 | Huruf latin p kecil |
| q | 71 | 113 | Huruf latin q kecil |
| r | 72 | 114 | Huruf latin r kecil |
| s | 73 | 115 | Huruf latin s kecil |
| t | 74 | 116 | Huruf latin t kecil |
| u | 75 | 117 | Huruf latin u kecil |
| v | 76 | 118 | Huruf latin v kecil |
| w | 77 | 119 | Huruf latin w kecil |
| x | 78 | 120 | Huruf latin x kecil |
| y | 79 | 121 | Huruf latin y kecil |
| z | 007A | 122 | Huruf latin z kecil |

Berikut contoh perhitungan *rolling hash* dengan nilai k=3, bisa dilihat di Tabel 4.4 berikut.

Tabel 4 4.4 Perhitungan *Rolling Hash*

|  |  |  |
| --- | --- | --- |
| **Keterangan** | **Data Contoh** | ***N-Gram*** |
| *Dataset* | implementasipengamananfile | imp mpl ple lem eme men ent nta tas asi sip ipe pen eng nga gam ama man ana nan anf nfi fil ile |
| Data Tes | aplikasipengamananfile | apl pli lik ika kas asi sip ipe pen eng nga gam ama man ana nan anf nfi fil ile |

Pada Tabel 4.5 dan Tabel 4.6 akan menjelaskan contoh perhitungan dari potongan kata, untuk mencari nilai *hash* pada *dataset* dan data tes*.* Berikut penjabaran perhitungan nilai *hash* pada *dataset,* bisa dilihat di Tabel 4.6 berikut.

Tabel 5 4.5 Perhitungan *Hash Dataset*

| **Kalimat** | **Desimal ASCII** | **Penjabaran** | **Hasil** |
| --- | --- | --- | --- |
|
| imp | i=105 m=109 p=112 | 105\*7^(3-1)+109\*7^(3-2)+112\*7^(3-3) | 6020 |
| mpl | m=109 p=112 l=108 | 109\*7^(3-1)+112\*7^(3-2)+108\*7^(3-3) | 6233 |
| ple | p=112 l=108 e=101 | 112\*7^(3-1)+108\*7^(3-2)+101\*7^(3-3) | 6345 |
| lem | l=108 e=101 m=109 | 108\*7^(3-1)+101\*7^(3-2)+109\*7^(3-3) | 6108 |
| eme | e=101 m=109 e=101 | 101\*7^(3-1)+109\*7^(3-2)+101\*7^(3-3) | 5813 |
| men | m=109 e=101 n=110 | 109\*7^(3-1)+101\*7^(3-2)+110\*7^(3-3) | 6158 |
| ent | e=101 n=110 t=116 | 101\*7^(3-1)+110\*7^(3-2)+116\*7^(3-3) | 5835 |
| nta | n=110 t=116 a=97 | 110\*7^(3-1)+116\*7^(3-2)+97\*7^(3-3) | 6299 |
| tas | t=116 a=97 s=115 | 116\*7^(3-1)+97\*7^(3-2)+115\*7^(3-3) | 6478 |
| asi | a=97 s=115 i=105 | 97\*7^(3-1)+115\*7^(3-2)+105\*7^(3-3) | 5663 |
| sip | s=115 i=105 p=112 | 115\*7^(3-1)+105\*7^(3-2)+112\*7^(3-3) | 6482 |
| ipe | i=105 p=112 e=101 | 105\*7^(3-1)+112\*7^(3-2)+101\*7^(3-3) | 6030 |
| pen | p=112 e=101 n=110 | 112\*7^(3-1)+101\*7^(3-2)+110\*7^(3-3) | 6305 |
| eng | e=101 n=110 g=103 | 101\*7^(3-1)+110\*7^(3-2)+103\*7^(3-3) | 5822 |
| nga | n=110 g=103 a=97 | 110\*7^(3-1)+103\*7^(3-2)+97\*7^(3-3) | 6208 |
| gam | g=103 a=97 m=109 | 103\*7^(3-1)+97\*7^(3-2)+109\*7^(3-3) | 5835 |
| ama | a=97 m=109 a=97 | 97\*7^(3-1)+109\*7^(3-2)+97\*7^(3-3) | 5613 |
| man | m=109 a=97 n=110 | 109\*7^(3-1)+97\*7^(3-2)+110\*7^(3-3) | 6130 |
| ana | a=97 n=110 a=97 | 97\*7^(3-1)+110\*7^(3-2)+97\*7^(3-3) | 5620 |
| nan | n=110 a=97 n=110 | 110\*7^(3-1)+97\*7^(3-2)+110\*7^(3-3) | 6179 |
| anf | a=97 n=110 f=102 | 97\*7^(3-1)+110\*7^(3-2)+102\*7^(3-3) | 5625 |
| nfi | n=110 f=102 i=105 | 110\*7^(3-1)+102\*7^(3-2)+105\*7^(3-3) | 6209 |
| fil | f=102 i=105 l=108 | 102\*7^(3-1)+105\*7^(3-2)+108\*7^(3-3) | 5841 |
| ile | i=105 l=108 e=101 | 105\*7^(3-1)+108\*7^(3-2)+101\*7^(3-3) | 6002 |

Setelah dilakukan perhitungan diperoleh hasil dari *rolling hash* “6020 6233 6345 6108 5813 6158 5835 6299 6478 5663 6482 6030 6305 5822 6208 5835 5613 6130 5620 6179 5625 6209 5841 6002”.

Berikut penjabaran perhitungan nilai *hash* pada data tes, bisa dilihat di Tabel 4.6 berikut.

Tabel 6 4.6 Perhitungan *Hash* Data Tes

| **Kalimat** | **Desimal ASCII** | **Penjabaran** | **Hasil** |
| --- | --- | --- | --- |
|
| apl | a=97 p=112 l=108 | 97\*7^(3-1)+112\*7^(3-2)+108\*7^(3-3) | 5645 |
| pli | p=112 l=108 i=105 | 112\*7^(3-1)+108\*7^(3-2)+105\*7^(3-3) | 6349 |
| lik | l=108 i=105 k=107 | 108\*7^(3-1)+105\*7^(3-2)+107\*7^(3-3) | 6134 |
| ika | i=105 k=107 a=97 | 105\*7^(3-1)+107\*7^(3-2)+97\*7^(3-3) | 5991 |
| kas | k=107 a=97 s=115 | 107\*7^(3-1)+97\*7^(3-2)+115\*7^(3-3) | 6037 |
| asi | a=97 s=115 i=105 | 97\*7^(3-1)+115\*7^(3-2)+105\*7^(3-3) | 5663 |
| sip | s=115 i=105 p=112 | 115\*7^(3-1)+105\*7^(3-2)+112\*7^(3-3) | 6482 |
| ipe | i=105 p=112 e=101 | 105\*7^(3-1)+112\*7^(3-2)+101\*7^(3-3) | 6030 |
| pen | p=112 e=101 n=110 | 112\*7^(3-1)+101\*7^(3-2)+110\*7^(3-3) | 6305 |
| eng | e=101 n=110 g=103 | 101\*7^(3-1)+110\*7^(3-2)+103\*7^(3-3) | 5822 |
| nga | n=110 g=103 a=97 | 110\*7^(3-1)+103\*7^(3-2)+97\*7^(3-3) | 6208 |
| gam | g=103 a=97 m=109 | 103\*7^(3-1)+97\*7^(3-2)+109\*7^(3-3) | 5835 |
| ama | a=97 m=109 a=97 | 97\*7^(3-1)+109\*7^(3-2)+97\*7^(3-3) | 5613 |
| man | m=109 a=97 n=110 | 109\*7^(3-1)+97\*7^(3-2)+110\*7^(3-3) | 6130 |
| ana | a=97 n=110 a=97 | 97\*7^(3-1)+110\*7^(3-2)+97\*7^(3-3) | 5620 |
| nan | n=110 a=97 n=110 | 110\*7^(3-1)+97\*7^(3-2)+110\*7^(3-3) | 6179 |
| anf | a=97 n=110 f=102 | 97\*7^(3-1)+110\*7^(3-2)+102\*7^(3-3) | 5625 |
| nfi | n=110 f=102 i=105 | 110\*7^(3-1)+102\*7^(3-2)+105\*7^(3-3) | 6209 |
| fil | f=102 i=105  l= 108 | 102\*7^(3-1)+105\*7^(3-2)+108\*7^(3-3) | 5841 |
| ile | i=105 l=108 e=101 | 105\*7^(3-1)+108\*7^(3-2)+101\*7^(3-3) | 6002 |

Setelah dilakukan perhitungan diperoleh hasil dari *rolling hash* “5645 6349 6134 5991 6037 5663 6482 6030 6305 5822 6208 5835 5613 6130 5620 6179 5625 6209 5841 6002”.

### Pembentukan *Window*

Setelah didapatkan hasil perhitungan *rolling hash,* maka tahap selanjutnya adalah pembentukan *window* dari nilai *hash* yang didapat. Kemudian nilai *hash* dikelompokan sebanyak nilai *w-gram* atau w=3. Berikut contoh pembentukan *window.*

1. *Dataset*

Berikut pembentukan *window* pada *dataset* dengan w=3.

Nilai *hash :*

6020 | 6233 | 6345 | 6108 | 5813 | 6158 | 5835 | 6299 | 6478 | 5663 | 6482 | 6030 | 6305 | 5822 | 6208 | 5835 | 5613 | 6130 | 5620 | 6179 | 5625 | 6209 | 5841 | 6002

*Window* (w=3) :

6020 | 6233 | 6345

6108 | 5813 | 6158

5835 | 6299 | 6478

5663 | 6482 | 6030

6305 | 5822 | 6208

5835 | 5613 | 6130

5620 | 6179 | 5625

6209 | 5841 | 6002

1. Data Tes

Berikut pembentukan *window* pada data tesdengan w=3.

Nilai *hash :*

5645 | 6349 | 6134 | 5991 | 6037 | 5663 | 6482 | 6030 | 6305 | 5822 | 6208 | 5835 | 5613 | 6130 | 5620 | 6179 | 5625 | 6209 | 5841 | 6002

*Window* (w=3) :

5645 | 6349 | 6134

5991 | 6037 | 5663

6482 | 6030 | 6305

5822 | 6208 | 5835

5613 | 6130 | 5620

6179 | 5625 | 6209

6209 | 5841 | 6002

### Pencarian *Fingerprint*

Setelah pembentukan *window* selesai, maka tahap selanjutnya adalah pencarian *fingerprint* dari setiap *window* yang ada. Nilai *fingerprint* ditentukan berdasarkan nilai terkecil dari setiap *window* yang ada. Berikut hasil dari *fingerprint* dari contoh sebelumnya.

1. *Dataset*

Berikut hasil pencarian *fingerprint* dari *dataset:*

6020 | 5813 | 5835 | 5663 | 5822 | 5613 | 5620 | 5841

1. Data Tes

Berikut hasil pencarian *fingerprint* dari data tes:

5645 | 5663 | 6030 | 5822 | 5613 | 5625 | 5841

### 4.2.5. *Jaccard Similarity*

Setelah *fingerprint* dari dokumen ditemukan langkah selanjutnya adalah perhitungan *similarity* dari *fingerprint* yang ada menggunakan *jaccard Similarity*. Berikut contoh perhitungan *jaccard similarity*.

Berikut hasil setiap *fingerprint* ada pada Tabel 4.7 berikut.

Tabel 7 4.7 Hasil *Fingerprint*

|  |  |  |
| --- | --- | --- |
| **Keterangan** | **Contoh Kalimat** | ***Fingerprint*** |
| *Dataset* | implementasipengamananfile | 6020 | 5813 | 5835 | 5663 | 5822 | 5613 | 5620 | 5841 |
| Data Tes | aplikasipengamananfile | 5645 | 5663 |6030 | 5822 | 5613 | 5625 | 5841 |

Berikut adalah rumus *jaccard similarity* :

*Similarity* (X,Y) =x 100%

X = {6020,5813,5835,5663,5822,5613,5620,5841}

Y = {5645,5663,6030,5822,5613,5625,5841}

X∩Y = {5663,5822,5613,5841}

X∪Y = {6020,5813,5835,5663,5822,5613,5620,5841,5645,6030,5625}

(X,Y) =x 100%

(X,Y) =x 100% = 0,363 x 100%= **36,3%**

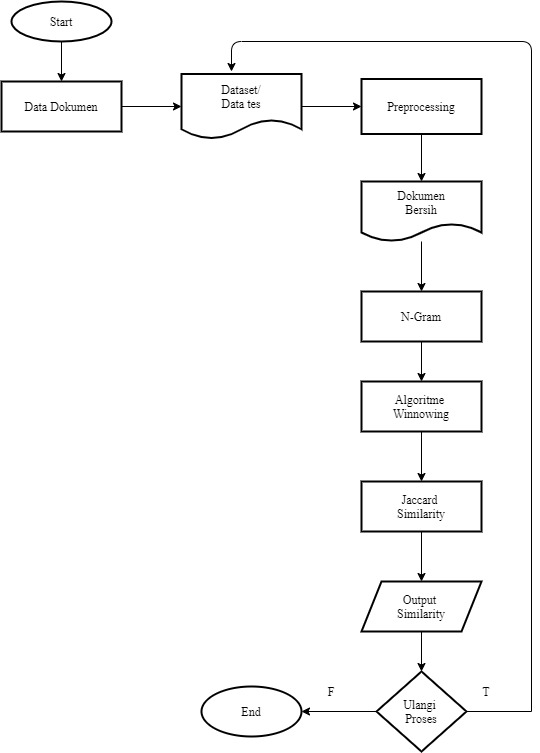
Sehingga tingkat *similarity* dari dokumenadalah 36,3%

## 4.3. *Flowchart* Tahapan Metode

*Flowchart* adalah suatu bagan atau simbol-simbol yang menggambarkan alur kerja atau urutan proses pada suatu program. Berikut adalah penjabaran *flowchart* pada tahapan metode yang digunakan:

### 4.3.1. *Flowchart* Keseluruhan Sistem

Pada *flowchart* ini menjelaskan tahapan mengenai berjalannya sistem, mulai dari pengumpulan data dokumen, hingga mendapatkan hasil tingkat *similarity* dari sebuah dokumen.

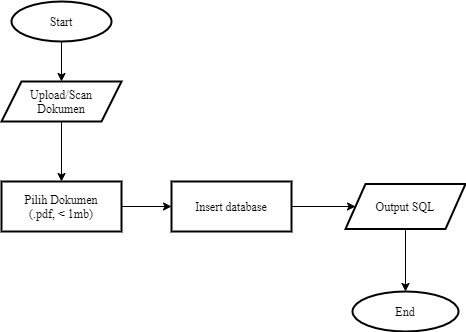


Gambar 1 4.1 *Flowchart* Keseluruhan Sistem

Pada gambar 4.1 menjelaskan proses keseluruhan sistem yang dibuat dengan tahap awal yaitu proses pengumpulan data dokumen, kemudian sebelum masuk ke tahapan n*-gram*, dokumen yang merupakan *dataset* atau data tes ini dilakukan tahapan pembersihan data atau *preprocessing*, setelah menjadi dokumen bersih, selanjutnya dilakukan proses tahapan utama *n-gram*, pembentukan *hash* melalui algoritme *winnowing,* perhitungan *similarity* dengan *jaccard similarity* hingga menghasilkan output, berupa tingkat *similarity*. Proses yang sama akan berulang jika dimasukkan dokumen untuk dijadikan *dataset* atau data tes.

### 4.3.2. *Flowchart* Data Dokumen

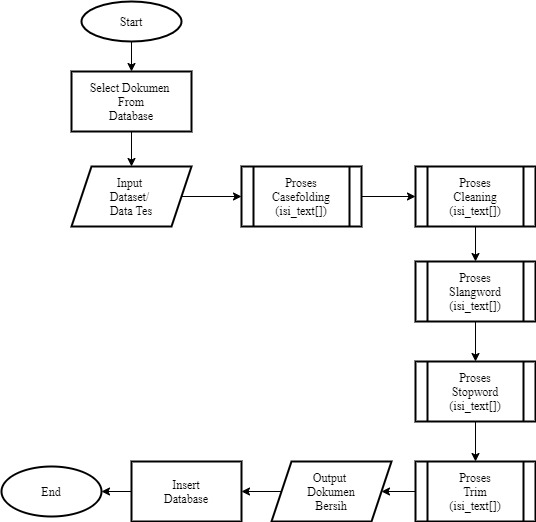
Pada *flowchart* ini, menjelaskan tahapan proses pengumpulan *dataset* serta proses *scan* untuk pengecekan tingkat *similarity* sebuah dokumen. Dokumen yang bisa diproses berupa file *.pdf* serta berukuran <1mb, setelah itu dokumen yang sudah diproses akan disimpan ke *database.* Proses yang sama akan berulang jika ingin memproses dokumen kembali.



Gambar 2 4.2 *Flowchart* Data Dokumen

### 4.3.3. *Flowchart Preprocessing*

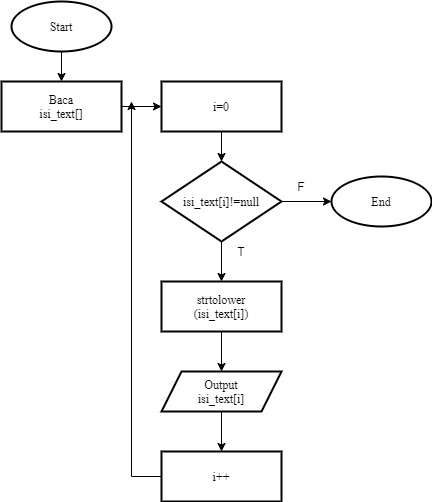
Pada *flowchart* ini, menjelaskan tahapan *preprocessing* data dokumen. Data dokumen hasil *upload/scan* yang sudah di*impor* ke dalam *database* dipilih untuk kemudian dilakukan sub proses *casefolding, cleaning*, *slangword, stopword,* dan penghapusan spasi antar kata. Setelah itu dokumen bersih disimpan ke dalam *array* isi\_text[], dan dimasukkan kembali ke dalam *database*.



Gambar 3 4.3 *Flowchart Preprocessing*

### 4.3.4. *Flowchart Casefolding*

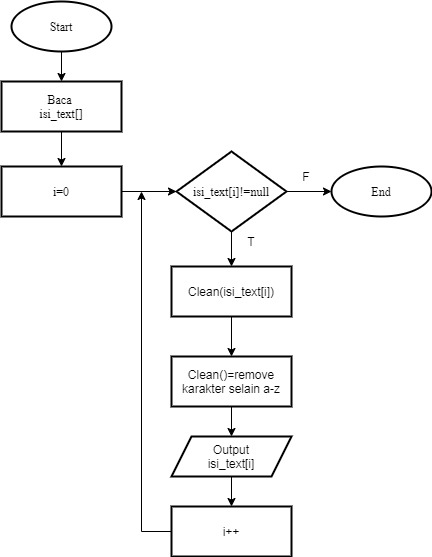
Pada *flowchart* ini, menjelaskan sub proses tahapan dari *preprocessing* yaitu *casefolding*. Pada tahapan ini, dokumen yang sudah tersimpan pada *array* isi\_text[] satu per satu dilakukan proses *casefolding* yaitu perubahan isi dokumen menjadi huruf kecil. Dengan fungsi strtolower(). Proses ini dijalankan sampai dokumen habis atau dokumen paling terakhir.



Gambar 4 4.4 *Flowchart Casefolding*

### 4.3.5. *Flowchart Cleaning*

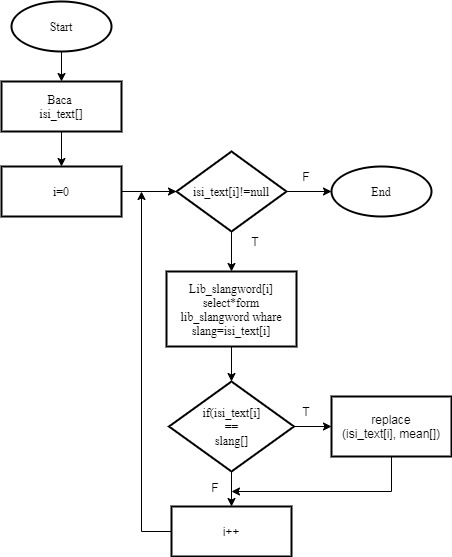
Pada *flowchart* ini, menjelaskan sub proses tahapan dari *preprocessing* yaitu *cleaning*. Pada tahapan ini, dokumen yang sudah tersimpan pada *array* isi\_text[] satu per satu dilakukan penghilangan pada karakter selain a-z dengan fungsi clean(). Proses ini dijalankan sampai dokumen habis atau dokumen paling terakhir.



Gambar 5 4.5 *Flowchart Cleaning*

### 4.3.6. *Flowchart Slangword*

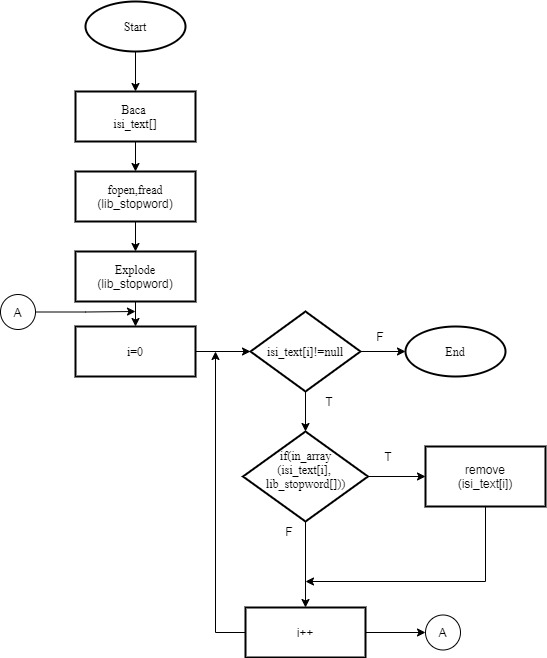
Pada *flowchart* ini, menjelaskan sub proses tahapan dari *preprocessing* yaitu *slangword.* Proses *slangword* diawali dengan membaca dokumen pada *array* isi\_text[] dan memanggil *library slangword* yang sama pada *array* isi\_text[i], jika ada yang sama maka dilakukan perintah replace(isi\_text[i], mean) yang artinya mengganti kata tersebut dengan arti pada kamus *library slangword*. Proses ini berulang sampai dokumen habis atau dokumen paling terakhir.



Gambar 6 4.6 *Flowchart Slangword*

### 4.3.7. *Flowchart Stopword*

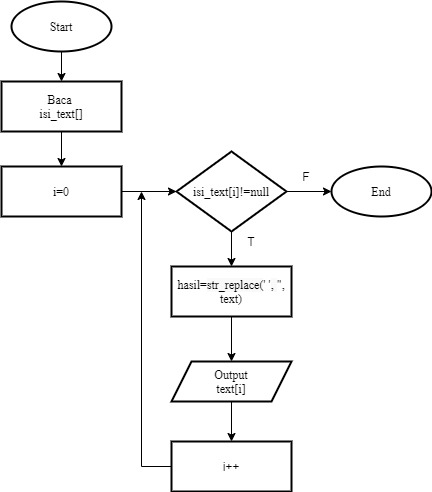
Pada *flowchart* ini, menjelaskan sub proses tahapan dari *preprocessing* yaitu *stopword,* proses ini diawali dengan membaca *library stopword* yang berisi kamus *stopword.* Kemudian dilakukan pencarian dengan fungsi in\_array (isi\_text[i], lib-stopword[]), artinya jika sebuah kata dalam array isi\_text[i] terdapat dalam kamus *stopword,* maka kata tersebut dihapus menggunakan fungsi remove(isi\_text[i]). Proses ini berulang sampai dokumen habis atau dokumen paling terakhir.



Gambar 7 4.7 *Flowchart Stopword*

### 4.3.8. *Flowchart* Hapus Spasi

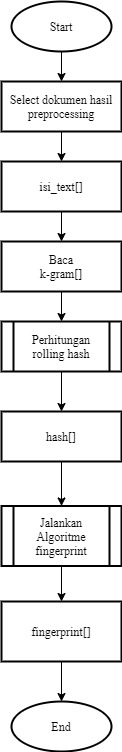
Pada *flowchart* ini, menjelaskan sub proses tahapan dari *preprocessing* yaitu hapus spasi. Pada tahapan ini, dokumen yang sudah tersimpan pada *array* isi\_text[] satu per satu dilakukan penghapusan spasi antara satu kata dengan kata lainya. Menggunakan fungsi str\_replace(). Proses ini berulang sampai dokumen habis atau dokumen paling terakhir.



Gambar 8 4.8 *Flowchart* Hapus Spasi

### 4.3.9. *Flowchart* Algoritme *Winnowing*

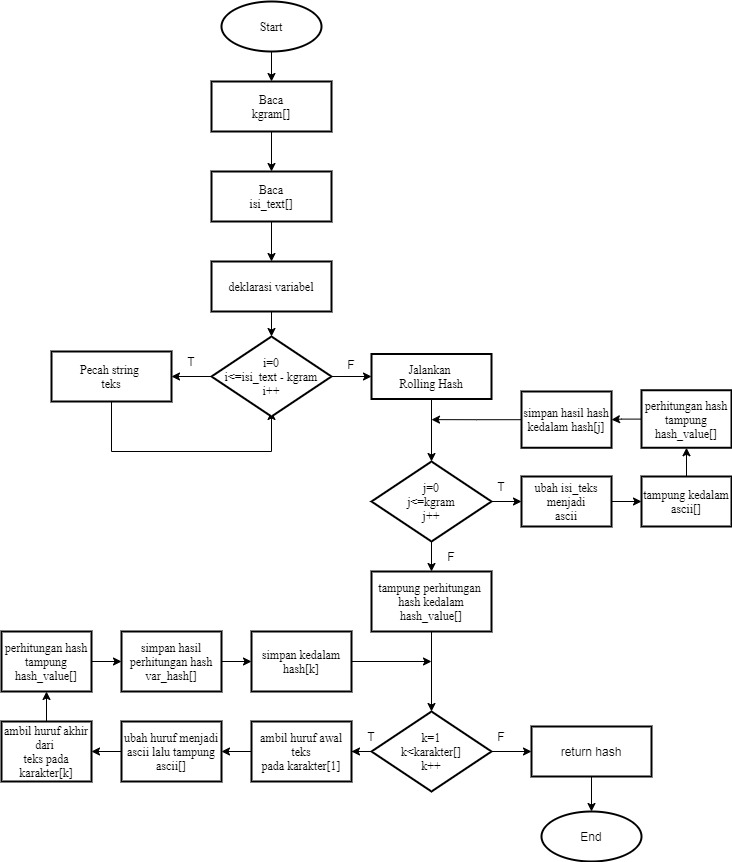
Pada *flowchart* ini, menjelaskan tahapan dari algoritme *winnowing* untuk mencari nilai *hash* dari sebuah dokumen. Setelah dokumen yang diproses pada tahap *preprocessing* dan telah disimpan kedalam *database,* kemudian dicarilah nilai *hash* dari dokumen tersebut, melalui beberapa sub proses *rolling hash,* serta pencarian *fingerprint.* Setelah mendapatkan nilai *hash* dari dokumen yang diproses, maka setelah itu dapat dilakukan proses pencarian tingkat *similarity* dari sebuah dokumen.



Gambar 9 4.9 *Flowchart* Algoritme *Winnowing*

### 4.3.10. *Flowchart Rolling Hash*

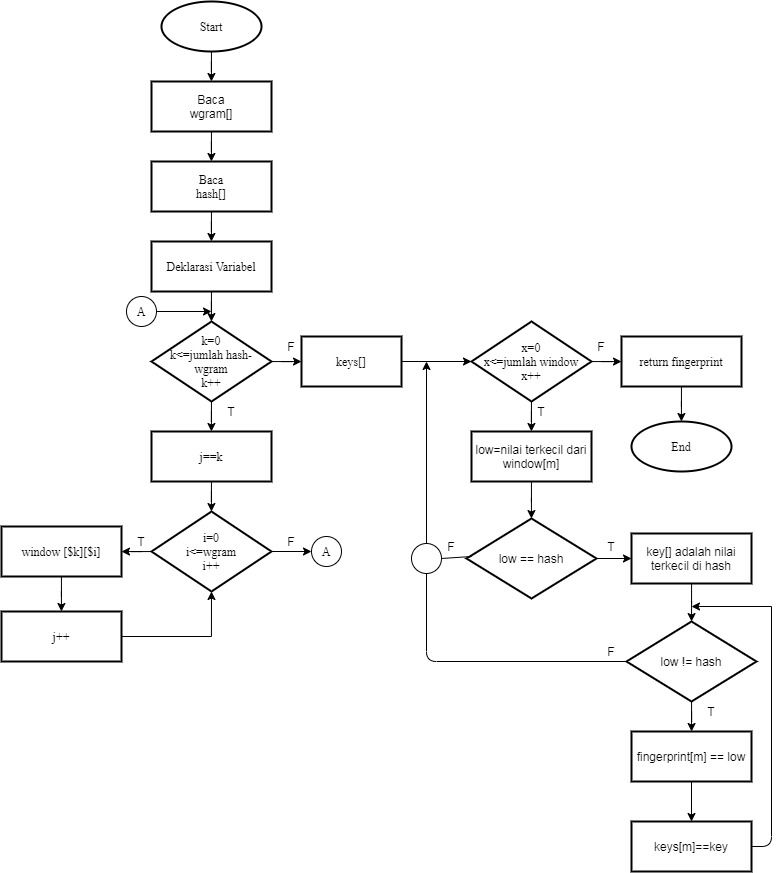
Pada *flowchart* ini, menjelaskan tahapan dari *rolling hash* yang merupakan sub proses dari algoritme *winnowing.* Proses pertama yang dilakukan adalah penentuan nilai *k-gram* untuk menentukan jumlah pecahan teks yang akan diproses, setelah teks dipecah, maka proses selanjutnya mencari nilai ASCII dari setiap huruf yang dipecah, kemudian ditampung di *variabel* hash\_value(), setelah mendapatkan nilai ASCII dari setiap huruf yang akan diproses, maka proses selanjutnya mencari nilai *hash* dari teks tersebut.



Gambar 10 4.10 *Flowchart Rolling Hash*

### 4.3.11. *Flowchart Fingerpring*

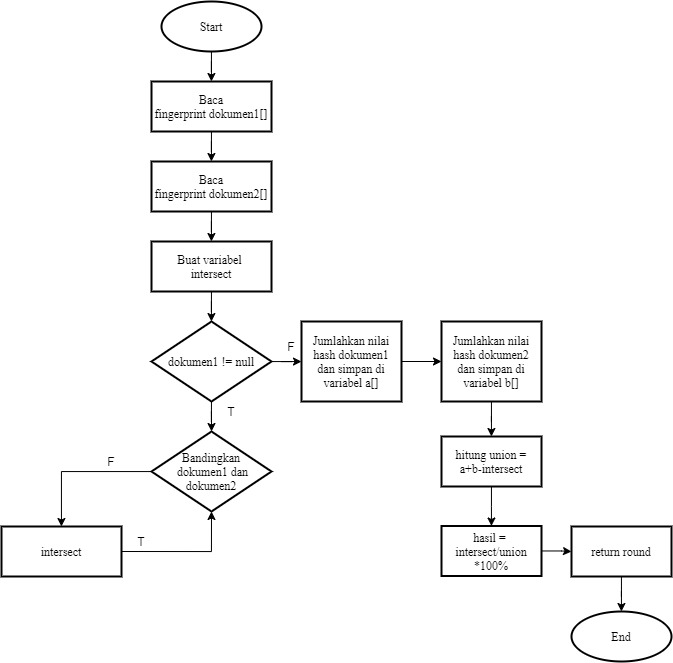
Pada *flowchart* ini, menjelaskan tahapan pencarian nilai *fingerprint* dari proses sebelumnya. Proses pertama yang dilakukan adalah pembentukan nilai *window,* pembentukan ini bertujuan untuk mengelompokan nilai *hash* yang telah didapatkan. Setelah dilakukan pembentukan *window,* maka tahap selenjutnya mencari nilai terkecil dari setiap *window* yang dibentuk, sehingga didapatkanlah nilai *fingerprint* dari setiap *window* yang telah dibentuk.



Gambar 11 4.11 *Flowchart Fingerpring*

### 4.3.12. *Flowchart Jaccard Similarity*

Pada *flowchart* ini, menjelaskan tahapan dari *jaccard similarity.* Tahapan ini merupakan tahapan untuk mencari persentase *similarity* dari sebuah dokumen. Setelah didapatkan nilai *fingerprint* pada tahap sebelumnya, maka pada proses selanjutnya adalah membandingkan nilai *fingerprint* dari dokumen *dataset* dengan dokumen data tes, sehingga dapat diperoleh persentase *similarity* dari dokumen yang telah diproses.



Gambar 12 4.12 *Flowchart Jaccard Similarity*

## 4.4. Algoritme Tahapan Metode

Algoritme adalah urutan atau alur tahapan proses yang dijabarkan dalam bentuk tulisan, algoritme ini merupakan representasi dari flowchart yang telah dijelaskan sebelumnya.

### 4.4.1. Algoritme Keseluruhan Sistem

Pada algoritme ini dijelaskan tentang proses sistem secara keseluruhan pada metode yang digunakan.

**Algoritme 4.1 Keseluruhan Sistem**

|  |
| --- |
|  |

### 4.4.2. Algoritme Data Dokumen

Pada algoritme ini dijelaskan tentang proses tahapan pengolahan dokumen yang akan diproses.

**Algoritme 4.2 Data Dokumen**

|  |
| --- |
|  |

### 4.4.3. Algoritme *Preprocessing*

Pada algoritme ini dijelaskan tentang proses tahapan pada *preprocessing* secara keseluruhan.

**Algoritme 4.3 *Preprocessing***

|  |
| --- |
|  |

### 4.4.4. Algoritme *Casefolding*

Pada algoritme ini dijelaskan tentang proses sub tahapan pada *preprocessing* yaitu *casefolding*.

**Algoritme 4.4 *Casefolding***

|  |
| --- |
|  |

### 4.4.5. Algoritme *Cleaning*

Pada algoritme ini dijelaskan tentang proses sub tahapan pada *preprocessing* yaitu *cleaning* dokumen.

**Algoritme 4.5 *Cleaning***

|  |
| --- |
|  |

### 4.4.6. Algoritme *Slangword*

Pada algoritme ini dijelaskan tentang proses sub tahapan pada *preprocessing* yaitu *slangword*.

**Algoritme 4.6 *Slangword***

|  |
| --- |
|  |

### 4.4.7. Algoritme *Stopword*

Pada algoritme ini dijelaskan tentang proses sub tahapan pada *preprocessing* yaitu *slangword*.

**Algoritme 4.7 *Stopword***

|  |
| --- |
|  |

### 4.4.8. Algoritme Hapus Spasi

Pada algoritme ini dijelaskan tentang proses sub tahapan pada *preprocessing* yaitu menghapus spasi pada sebuah teks dokumen.

**Algoritme 4.8 Hapus Spasi**

|  |
| --- |
|  |

### 4.4.9. Algoritme *Winnowing*

Pada algoritme ini dijelaskan tentang proses tahapan utama pada algoritme *winnowing* yaitu *rolling hash* dan  *fingerprint.*

**Algoritme 4.9 *Winnowing***

|  |
| --- |
|  |

### 4.4.10. Algoritme *Rolling Hash*

Pada algoritme ini dijelaskan tentang proses sub tahapan pada algoritme *winnowing* yaitu *rolling hash*.

**Algoritme 4.10 *Rolling Hash***

|  |
| --- |
|  |

### 4.4.11. Algoritme *Fingerprint*

Pada algoritme ini dijelaskan tentang proses sub tahapan pada algoritme *winnowing* yaitu *rolling hash*.

**Algoritme 4.11 *Fingerprint***

|  |
| --- |
|  |

### 4.4.12. Algoritme *Jaccard Similarity*

Pada algoritme ini dijelaskan tentang proses tahapan metode *jaccard similarity.*

**Algoritme 4.12 *Jaccard Similarity***

|  |
| --- |
|  |

## 4.5. Pengujian

## 4.6. Tampilan Layar Aplikasi