

## **BAB V**

### **IMPLEMENTASI DAN PENGUJIAN**

#### **5.1. Tahapan Implementasi**

Tahap implementasi ini dilakukan setelah tahap analisa dan perancangan selesai dilakukan. Dalam tahap implementasi ini akan dilakukan pengkodean terhadap apa yang telah dilakukan pada tahap analisa dan perancangan. Diharapkan hasil dari implementasi ini sesuai dengan apa yang menjadi tujuan awal penelitian ini.

##### **5.1.1. Batasan Implementasi**

Adapun batasan implementasi pada aplikasi dokumen clustering ini adalah sebagai berikut:

1. Input masukan dokumen teks yang akan di cluster memiliki ekstensi *Plaintext* (\*.txt), ini dimaksudkan karena proses dalam konversi dokumen teks yang panjang dan membutuhkan waktu yang lama.
2. Dokumen yang di uji dalam proses dokumen clustering ini akan di batasi sebanyak 20 dokumen, karena proses pengujian menggunakan dua sistem metode yaitu untuk *jaccard coefficient* dan *euclidian distance* sehingga akan memberikan waktu efektifitas kinerja yang lebih lama dan juga untuk lebih memudahkan dalam proses pengelompokan didalam laporan pengujian.

##### **5.1.2. Lingkungan Implementasi**

Lingkungan implementasi adalah lingkungan dimana aplikasi ini dikembangkan. Lingkungan implementasi aplikasi dokumen clustering ini terdiri atas dua lingkungan yaitu lingkungan perangkat keras dan perangkat lunak.

Berikut ini merupakan spesifikasi lingkungan tersebut:

1. Perangkat keras

Berikut spesifikasi dari perangkat keras yang digunakan:

Processor : *AMD A8-4500M APU Radeon™ 1.90 Ghz*

Memori (RAM) : 4,00 GB

Harddisk : 500 GB

2. Perangkat Lunak

Sistem Operasi : *Windows 7*

Bahasa Pemrograman : *PHP (Hypertext Preprocessor)*

*Tools* Perancangan : *Notepad++, Google Chrome*

### **5.1.3. Implementasi Antarmuka Aplikasi**

Tahapan ini merupakan tahap hasil implementasi dokumen clustering dari hasil analisa pengkodean yang telah dilakukan dan di implementasikan ke dalam bentuk *interface*(antarmuka) yang telah dibuat. Pada tahapan *interface* ini aplikasi akan terdiri dari enam menu utama yaitu: *Home* (beranda), *Corpus*, *Setting*, *Clustering*, *List Uji*, dan *About*.

#### **1. Halaman Menu *Home* (Beranda)**

Menu *Home* (beranda) merupakan tampilan dari halaman yang berisi tentang definisi singkat mengenai Algoritma Winnowing dan Algoritma K-Means yang menjadi metode dalam pembuatan aplikasi dokumen clustering.

Berikut adalah tampilan menu *home*(beranda):



Gambar 5.1 Tampilan Menu Home

## 2. Halaman Menu Corpus Dokumen

Menu Corpus Dokumen merupakan tampilan dari halaman aplikasi dokumen clustering yang berfungsi untuk menampilkan daftar dokumen yang telah kita input, di dalam direktori *local c:/xampp/htdocs/dokumen clustering1/data/file*. Untuk file dokumen yang akan diuji untuk sistem dengan *Euclidian distance* dapat menginputkan di dokumen clustering 2.



Gambar 5.2 Tampilan Menu Corpus Dokumen

### 3. Halaman Menu Setting

Menu Setting merupakan halaman yang menampilkan fungsi untuk penginputan proses algoritma *winnowing* dan pemberian nilai kelompok yang akan di proses. Prosedur Menu setting ini terdiri dari input kode uji, input nilai k-gram, input nilai basis bilangan, input window, dan kelompok.

Berikut adalah tampilan interface Menu *Setting*:



The screenshot shows the 'DOCUMENT CLUSTERING' web application interface. The header features the logo of UIN SUSKA and navigation links: Home, Corpus Doc, Setting (highlighted), Clustering, List Uji, and About. The main content area is divided into two sections. On the left, there is an 'Info' section with a clock icon. On the right, there is a form with the following fields and values: 'Kode Uji' (8), 'K-Gram' (8), 'Basis Bilangan Prima' (3), 'Window' (8), and 'Kelompok' (4). A 'Simpan' button is located at the bottom right of the form.

Gambar 5.3 Tampilan Menu *Setting*

### 4. Halaman Menu *Clustering*

Menu *clustering* merupakan menu yang menampilkan halaman proses *cluster k-means*, berupa informasi mengenai dokumen teks yang telah di klaster, dan dikelompokkan berdasarkan anggota dokumen, dan menampilkan nilai *similarity* hasil *clustering*. Di menu ini juga terdapat menu tombol tampilkan proses, tombol cek gram, dan file cluster.

Berikut ini adalah tampilan interface menu *Clustering*:



Gambar 5.4 TampilanMenu *Clustering Jaccard Coefficient*



Gambar 5.5 TampilanMenu *Clustering Euclidian Distance*

## 5. Halaman Menu Tampilkan Proses

Pada form tampilan Proses ini akan menampilkan informasi proses setelah *clustering* dilakukan, dimana akan menampilkan halaman mengenai proses seluruh iterasi yang didapat setelah proses clustering sesuai dengan nilai *similarity* masing – masing dokumen.

The screenshot shows the 'DOCUMENT CLUSTERING' application interface. The top navigation bar includes 'Home', 'Corpus Doc', 'Setting', 'Clustering' (highlighted), 'List Uji', and 'About'. On the left, there is an 'Info' section with a clock icon. The main content area displays three tables, each representing a different iteration of the clustering process. Each table has columns for 'No', 'Nama Dokumen', 'Kelompok', and 'Jaccard Coefficient'.

No	Nama Dokumen	Kelompok	Jaccard Coefficient
1	01.01.01.01.01	3	0.291022480017
2	02.01.01.01.01	3	0.280329744384
3	03.01.01.01.01	1	0.370444910233
4	04.01.01.01.01	1	0.295222894478
5	05.01.01.01.01	3	0.198923589208
6	06.01.01.01.01	3	0.188118138218
7	07.01.01.01.01	4	0.181671520638
8	08.01.01.01.01	3	0.148883888119

No	Nama Dokumen	Kelompok	Jaccard Coefficient
1	01.01.01.01.01	3	0.283804178169
2	02.01.01.01.01	3	0.2887214344787
3	03.01.01.01.01	1	0.180000000000
4	04.01.01.01.01	1	0.181782030583
5	05.01.01.01.01	2	0.213483168742
6	06.01.01.01.01	2	0.1
7	07.01.01.01.01	4	0.1
8	08.01.01.01.01	4	0.182240781277

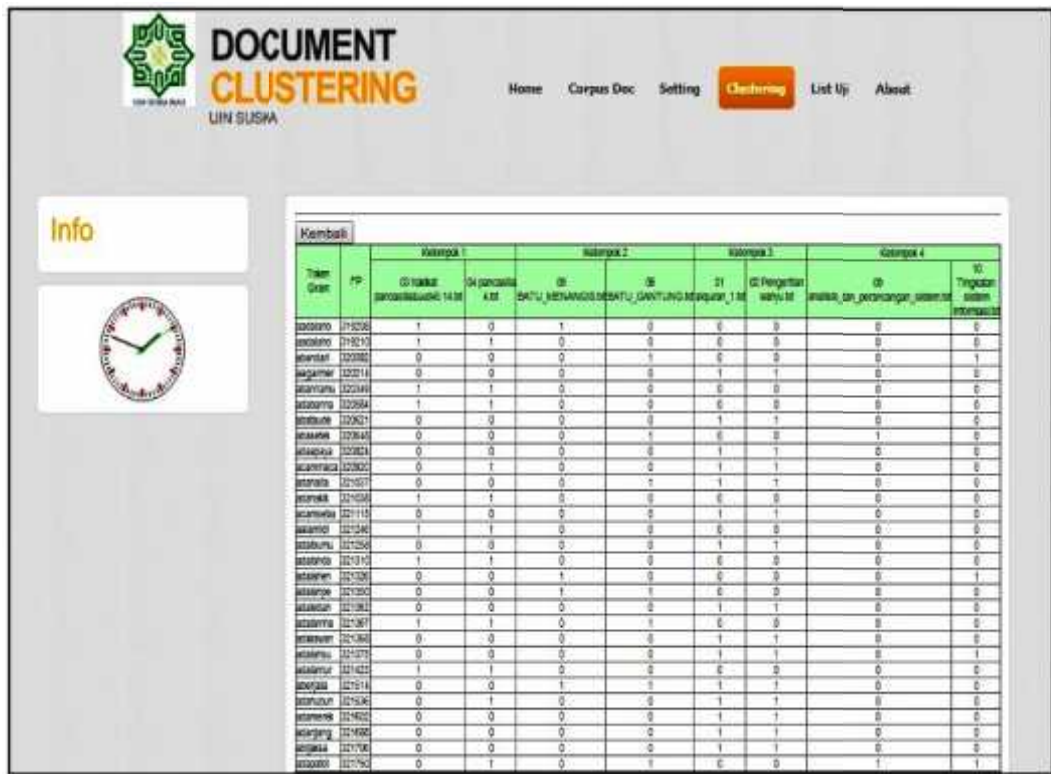
  

No	Nama Dokumen	Kelompok	Jaccard Coefficient
1	01.01.01.01.01	3	0.184197038859
2	02.01.01.01.01	3	0.008023587408
3	03.01.01.01.01	1	0.180000000000
4	04.01.01.01.01	1	0.181782030583
5	05.01.01.01.01	2	0.1481214082387
6	06.01.01.01.01	2	0.110000000000
7	07.01.01.01.01	4	0.1847712743223
8	08.01.01.01.01	4	0.138410841044

Gambar 5.6 Tampilan Menu Tampilkan Proses Iterasi

## 6. Halaman Menu Cek Gram

Pada form tampilan cek gram ini akan menampilkan informasi proses setelah clustering dilakukan, dimana akan menampilkan halaman mengenai Token Gram, *fingerprint*, matrik frekuensi anggota dokumen terhadap kelompoknya.



Gambar 5.7 Tampilan Menu Button Cek Gram

Pada menu ini juga sistem dapat melihat token gram, dan fingerprint yang relevan dengan masing – masing dokumen pada setiap kelompoknya.

Kembali		Kelompok 1		Kelompok 2		Kelompok 3		Kelompok 4	
Token Gram	FP	03 hakikat pancasila 14 bt	04 pancasila 4 bt	05 BATU_MENANGIS bt	06 BATU_GANTUNG bt	01 alquran_1 bt	02 Pengantar wahyu bt	09 analisis_dan_perancangan_sistem bt	10 Tingkatan sistem informasi bt
adalahb	319208	1	0	1	0	0	0	0	0
adalahd	319210	1	1	0	0	0	0	0	0
abandari	320082	0	0	0	1	0	0	0	1
agamer	320214	0	0	0	0	1	1	0	0
abannamu	320349	1	1	0	0	0	0	0	0
adabanna	320584	1	1	0	0	0	0	0	0
abetaude	320821	0	0	0	0	1	1	0	0
abasetele	320848	0	0	0	1	0	0	1	0
adaapaya	320824	0	0	0	0	1	1	0	0
acammaca	320929	0	1	0	0	1	1	0	0
adahajia	321037	0	0	0	1	1	1	0	0
adahakik	321038	1	1	0	0	0	0	0	0

Gambar 5.8 Tampilan Cek Fingerprint Relevan

## 7. Halaman Menu File Cluster

Pada halaman ini akan menampilkan partisi tempat penyimpanan hasil proses cluster yang telah dilakukan dan disimpan.

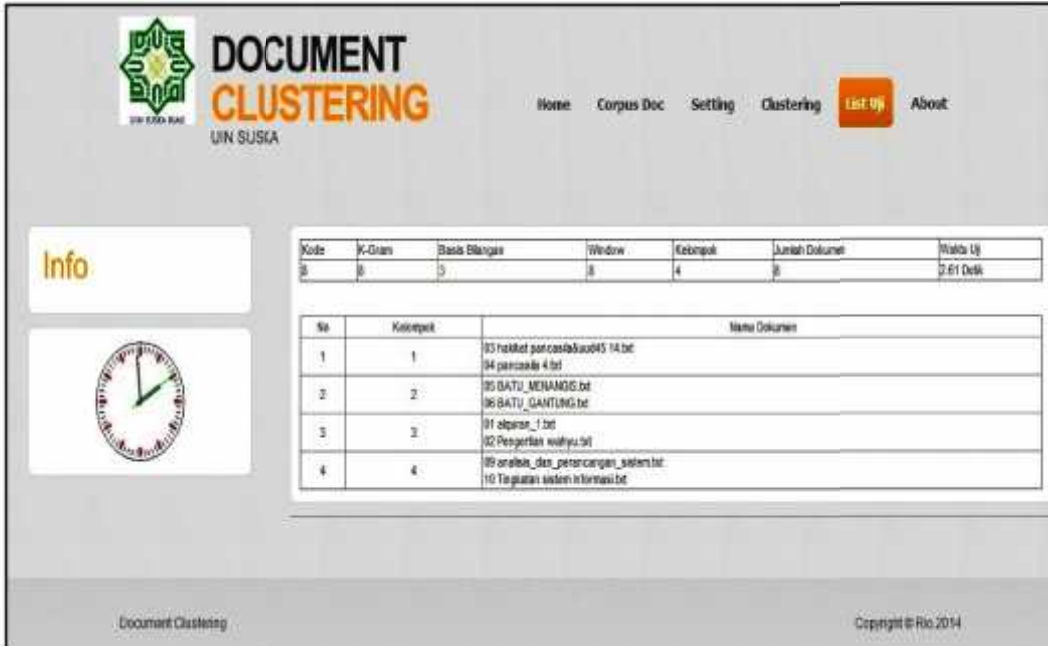


Name	Last modified	Size	Description
Parent Directory		-	
kelompok 1/	28-Apr-2014 07:25	-	
kelompok 2/	28-Apr-2014 07:25	-	
kelompok 3/	28-Apr-2014 07:25	-	
kelompok 4/	28-Apr-2014 07:25	-	

Gambar 5.9 Tampilan Menu Button File Cluster

## 8. Halaman Menu List Uji

Pada halaman tampilan menu list uji akan menampilkan informasi mengenai kode uji, k-gram, basis bilangan prima, window, kelompok, jumlah dokumen uji, waktu proses clustering yang dilakukan, dan menampilkan informasi hasil pengelompokan.



Kode	K-Gram	Basis Bilangan	Window	Kelompok	Jumlah Dokumen	Waktu Uji
8	8	3	8	4	8	2.61 Detik

No	Kelompok	Nama Dokumen
1	1	83 halikat_pancasila045 14.txt 84 pancasila 4.txt
2	2	85 BATU_MEMANGIS.txt 86 BATU_GANTUNG.txt
3	3	87 alquran_1.txt 88 Pengantar walhyu.txt
4	4	89 orasi_klas_penerangan_salem.txt 90 Tagihan sistem informasi.txt

Gambar 5.10 Tampilan Menu List Uji



## 9. Tampilan Form Menu About

Pada halaman ini akan menampilkan informasi mengenai penulis aplikasi dokumen clustering, pada halaman ini juga akan menampilkan informasi petunjuk penggunaan aplikasi dokumen clustering.



Gambar 5.11 Tampilan Menu About

## 5.2. Pengujian aplikasi

Pengujian sistem akan dilakukan untuk mengetahui apakah sistem yang dibangun sudah sesuai dengan analisa dan tujuan dari dibangunnya sistem ini. Untuk mengetahui hasil dari sistem ini apakah sudah sesuai dengan yang diharapkan, dilakukan penghitungan *similarity* dengan inputan dokumen yang diuji.

### 5.2.1 Rencana Pengujian

Adapun rencana pengujian yang akan dilakukan sebagai berikut:

1. Menguji nilai inputan awal pengujian berupa nilai  $k\text{-gram} = 8$ , basis bilangan  $= 3$ , window  $= 8$  sesuai inputan (Muhammad ridho, 2013). Untuk pengujian sendiri akan dibagi kedalam 2 pengujian yaitu pengujian I terdiri dari 4 tema yang didalamnya terdapat 3 subtema, dan pengujian II terdiri dari 5 tema.

2. Pengujian hasil running yang mengalami perubahan anggota cluster untuk mendapatkan nilai *precision* (ketepatan) dan nilai *relevansi* (rata-rata percobaan).
3. Pengujian *Blackbox*, yaitu untuk menguji program apakah telah sesuai dengan tujuan dan fungsi dan hasil yang didapat.

Pada pengujian ini dilakukan pengujian terhadap aplikasi dokumen clustering, dokumen input yang digunakan dalam scenario pengujian ini terdiri atas dokumen pengujian I dan dokumen pengujian II. Berikut penjelasan pengujian:

#### A. Pengujian I

Tahap pengujian I dilakukan dengan melakukan tes terhadap 20 dokumen input pada tabel:

**Tabel 5.1 Pengujian I Dokumen text**

No.	Tema	Subtema	Jumlah
1.	Pancasila	-	4
2.	Agama	Wahyu	3
		Tauhid	2
3.	Komputer	Jarkom	4
		Sistem Informasi	2
4.	Dongeng Anak	Cerita Rakyat	3
		Fabel	2
Jumlah: 20 Dokumen			

No	Daftar Dokumen
1	01 kunci-dan-buaya.txt
2	02 kunci-pak-tani.txt
3	03 ASAL_MULA_NAGARI_MINANGKABAU.txt
4	04 ASAL_MULA_TELAGA_BIRU.txt
5	05 BATANG_TUAKA.txt
6	06 fungsi-pancasila-sbg-dasar-negara.txt
7	07 butir-butir-pancasila.txt
8	08 hakikat_pancasila_sbg_dasar_negara.txt
9	09 pancasila_sbg_ideologi_terbuka.txt
10	10 istilah_wahyu.txt
11	11 pengertian_wahyu.txt
12	12 wahyu-dalam-islam.txt
13	13 belajar-tauhid-dan-akidah.txt
14	14 definisi-tauhid-dan-tauhid.txt
15	15 definisi_dan_merfaat_jaringan_komputer.txt
16	16 jaringan_komputer_nurut_para_ahli.txt
17	17 jenis-jenis_jarkom_dan_topologi.txt
18	18 konsep_jaringan_komputer.txt
19	19 analisis_dan_perancangan_sistem.txt
20	20 Konsep_dasar_sistem.txt

**Gambar 5.12 Daftar Dokumen Pengujian I**

## B. Pengujian II

Tahap pengujian II dilakukan dengan melakukan tes terhadap 20 dokumen teks, dimana pada dokumen ini memiliki *similarity* yang lebih baik dari tabel pengujian I, berikut input pengujian II pada tabel dibawah ini:

**Tabel 5.2 Pengujian II Dokumen Teks**

No.	Tema	Jumlah
1.	Pancasila	5
2.	Agama	3
3.	Dongeng Anak	4
4.	Jarkom	3
5	Sistem Informasi	5
Jumlah: 20 Dokumen		

No	Daftar Dokumen
1	01 alquran_1.txt
2	02 Pengertian wahyu.txt
3	03 ajaran_alquran_23.txt
4	05 pancasila 4.txt
5	06 manfaat pancasila 9.txt
6	07 hakikat pancasilaSuudi14.txt
7	08 piramida_pancasila 23.txt
8	09 sila-sila pancasila 6.txt
9	10 BATANG_TUAKA.txt
10	11 BATU_BATANGKUP.txt
11	12 BATU_MENANGIS.txt
12	13 BATU_GANTUNG.txt
13	13 jarkom.txt
14	14 Topologi jaringan.txt
15	15 Pembagian jarkom_12.txt
16	16 sisteminformasi.txt
17	17 Karakteristik_si.txt
18	18 pembagian prinsip_si_10.txt
19	19 si_19.txt
20	20 Tingkatan sistem informasi.txt

Gambar 5.13 Daftar Dokumen Pengujian II

### 5.2.1.1 Pengujian Nilai Inputan Program

Pengujian yang dilakukan dengan menggunakan dokumen input ini terdiri atas 20 dokumen yang memiliki 4 tema, dan 3 tema di antaranya memiliki subtema. Berikut informasi mengenai pengujian yang dilakukan:

- K-gram : 8
- Basis Bilangan : 3
- Window : 8
- Kelompok : 4

Pada konfigurasi ini akan dilakukan pengujian terhadap 20 dokumen yang berbeda isi dokumen tetapi memiliki pembahasan topik yang sama, sedangkan untuk konfigurasi akan menggunakan konfigurasi diatas.

Tabel 5.3 Pengujian I

No.	kelompok	Nama Dokumen	Keterangan
1	1	06 fungsi-pancasila-sbg-dasar-negara.txt	Relevan
		07 butir-butir-pancasila.txt	Relevan
		08 hakikat_pancasila_sbg_dasar_negara.txt	Relevan
		09 pancasila_sbg-ideologi_terbuka.txt	Relevan
		<b>Mayoritas Tema: Pancasila</b>	<b>Jumlah: 4</b>

2	2	01 kancil-dan-buaya.txt	Relevan
		02 kancil-pak-tani.txt	Relevan
		03 ASAL_MULA_NAGARI_MINANGKABAU.txt	Relevan
		05 BATANG_TUAKA.txt	Relevan
		13 belajar-tauhid-dari-alfatihah.txt	Tidak Relevan
		<b>Mayoritas Tema: Cerita Rakyat</b>	
3	3	04 ASAL_MULA_TELAGA_BIRU.txt	Tidak Relevan
		12 wahyu-dalam-islam.txt	Tidak Relevan
		14 defenisi-tauhid-dan-ilmu-tauhid.txt	Tidak Relevan
		15 defenisi_dan_manfaat_jaringan_komputer.txt	Relevan
		17 jenis-jenis_jarkom_dan_topologi.txt	Relevan
		18 konsep_jaringan_komputer.txt	Relevan
		19 analisis_dan_perancangan_sistem.txt	Relevan
		20 Konsep_dasar_sistem.txt	Relevan
		<b>Mayoritas Tema: Komputer</b>	
4	4	10 istilah_wahyu.txt	Relevan
		11 pengertian_wahyu.txt	Relevan
		16 jaringan_komputer_menurut_para_ahli.txt	Tidak Relevan
		<b>Mayoritas Tema: Agama</b>	

No	Nama Dokumen	Kelompok	Jaccard Coefficient
1	01 kancil-dan-buaya.txt	2	0.08993570171306
2	02 kancil-pak-tani.txt	2	0.066413962229009
3	03 ASAL_MULA_NAGARI_MINANGKABAU.txt	2	0.061135371179039
4	04 ASAL_MULA_TELAGA_BIRU.txt	3	0.055670103082784
5	05 BATANG_TUAKA.txt	2	0.09375
6	06 fungsi-pancasila-sbg-dasar-negara.txt	1	0.4010889292196
7	07 butir-butir-pancasila.txt	1	0.2944903187702
8	08 hakikat-pancasila-sbg-dasar-nejara.txt	1	0.35835351089588
9	09 pancasila-sbg-ideologi-terbuka.txt	1	0.35590277777776
10	10 istilah_wahyu.txt	4	0.182156133829
11	11 pengertian_wahyu.txt	4	0.2361518034884
12	12 wahyu-dalam-islam.txt	3	0.06257822779473
13	13 belajar-tauhid-dari-alfatihah.txt	2	0.065099457504521
14	14 defenisi-tauhid-dan-ilmu-tauhid.txt	3	0.053217321782178
15	15 defenisi_dan_manfaat_jaringan_komputer.txt	3	0.11201296701299
16	16 jaringan_komputer_menurut_para_ahli.txt	4	0.12807831773399
17	17 jenis-jenis_jarkom_dan_topologi.txt	3	0.097560975609756
18	18 konsep_jaringan_komputer.txt	3	0.10414201183432
19	19 analisis_dan_perancangan_sistem.txt	3	0.10297482837529
20	20 Konsep_dasar_sistem.txt	3	0.1159835542189

Kode	K-Gram	Basis Bilangan	Window	Kelompok	Jumlah Dokumen	Waktu Uji
18	8	3	8	4	20	8.83 Detik

Gambar 5.14 Screen-shoot Pengujian I Konfigurasi Jaccard

Dari tabel 5.3 diatas, terlihat konfigurasi nilai parameter input untuk  $k=8$ ,  $b=3$  dan  $w=8$ , pada hasil pengujian inputan awal didapatkan hasil sesuai dengan 4 kelompok yang akan dibentuk dimana terdapat beberapa hasil kelompok yaitu:

1. Pada kelompok 1 didapat 4 dokumen teks dengan mayoritas tema adalah pancasila dan hasil 4 dokumen relevan.
2. Pada kelompok 2 didapat 5 dokumen teks dengan mayoritas tema adalah cerita rakyat dan hasil 4 dokumen relevan.
3. Pada kelompok 3 didapat 8 dokumen teks dengan mayoritas tema adalah computer dan hasil 5 dokumen relevan.
4. Pada kelompok 4 didapat 3 dokumen teks dengan mayoritas tema adalah agama dan hasil 2 dokumen relevan.

### 5.2.1.2 Pengujian Hasil Running Program

Pengujian ini dilakukan untuk mengetahui anggota cluster yang mengalami perubahan cluster yang disebabkan pada algoritma k-means proses untuk membangkitkan centroid awal adalah acak (random).

#### 1. Dokumen teks 20 dengan 4 Kelompok (*Jaccard Coefficient*)

Pada pengujian ini diinputkan jumlah dokumen sebanyak 20 dengan isi setiap dokumen teks berbeda-beda dan dikelompokkan sebanyak 4 klaster. Menghitung nilai *precision* jumlah dokumen relevan yang didapat sistem dengan menggunakan persamaan 2.10 pada bab II, dan menghitung nilai relevansi dari *precision* menggunakan persamaan 2.11 pada bab II. Berikut hasil pengujian:

**Tabel 5.4** Rekapitulasi Hasil Running jaccardcoefficient 4

No.	Jaccard Coefficient								Waktu (s)	Precision (%)
	Kelompok Yang Relevan				Kelompok Yang Tidak Relevan					
	1	2	3	4	1	2	3	4		
<b>I</b>	4	6	4	3	2	0	1	0	12.16	85
<b>II</b>	5	3	2	4	1	1	2	3	16.44	70
<b>III</b>	3	2	3	6	3	1	2	0	16.44	70
<b>IV</b>	3	5	3	4	2	0	2	1	9.18	75

<b>V</b>	2	3	4	3	3	2	1	2	7.99	60
<b>VI</b>	3	5	4	3	0	1	1	3	11.29	75
<b>VII</b>	6	3	4	3	1	2	0	1	11.18	80
<b>VIII</b>	4	5	4	2	2	0	2	1	10.88	75
<b>IX</b>	2	4	5	4	2	2	0	1	14.20	75
<b>X</b>	5	2	4	3	1	1	2	2	9.56	70
<b>Relevansi</b>									<b>73.5</b>	

## 2. Dokumen teks 20 dengan 7 Kelompok (*Jaccard Coefficient*)

Pada pengujian ini diinputkan jumlah dokumen sebanyak 20 dengan isi setiap dokumen teks berbeda-beda dan dikelompokkan sebanyak 7 klaster.

Berikut hasil pengujian:

**Tabel 5.5** Rekapitulasi Hasil Running Jaccard Coefficient 7

No.	Jaccard Coefficient														Waktu (s)	Precision (%)
	Kelompok Yang Relevan							Kelompok Yang Tidak relevan								
	1	2	3	4	5	6	7	1	2	3	4	5	6	7		
<b>I</b>	2	2	0	1	4	3	2	1	1	3	0	0	0	1	22.86	70
<b>II</b>	3	2	3	2	3	1	1	0	1	1	1	0	0	2	15.35	75
<b>III</b>	1	4	1	2	4	4	3	1	0	0	0	0	0	0	10.96	95
<b>IV</b>	2	2	2	2	4	4	3	1	0	0	0	0	0	0	11.87	95
<b>V</b>	2	3	1	2	2	3	4	1	0	0	1	0	1	0	14.17	85
<b>VI</b>	2	4	2	2	1	1	4	1	1	0	2	0	0	0	9.46	80
<b>VII</b>	3	3	1	2	2	1	2	0	1	1	1	1	1	1	19.95	70
<b>VIII</b>	2	1	4	2	2	3	1	0	2	0	1	1	0	1	19.95	75
<b>IX</b>	3	1	2	3	2	2	2	1	0	1	1	1	1	0	18.06	75
<b>X</b>	2	1	1	2	1	4	3	1	0	0	2	1	1	1	28.46	70
<b>Relevansi</b>															<b>79</b>	

**3. Dokumen teks 20 dengan 5 Kelompok(Jaccard Coefficient)**

Pada pengujian ini diinputkan jumlah dokumen sebanyak 20 dengan isi setiap dokumen teks telah diatur tingkat similarity nya mencapai 70%, dan dikelompokkan sebanyak 5 klaster. Berikut hasil pengujian :

**Tabel 5.6 Rekapitulasi Hasil Running Jaccard Coefficient 5**

No.	Jaccard Coefficient										Waktu (s)	Precision (%)
	Kelompok Yang Relevan					Kelompok Yang Tidak Relevan						
	1	2	3	4	5	1	2	3	4	5		
<b>I</b>	4	3	5	5	3	0	0	0	0	0	5.83	100
<b>II</b>	3	5	5	3	4	0	0	0	0	0	6.73	100
<b>III</b>	3	5	3	4	3	0	0	0	2	0	7.40	90
<b>Relevansi</b>											<b>96.66</b>	

**4. Dokumen Teks 20 dengan 4 Kelompok(Euclidian Distance)**

Pada pengujian ini diinputkan jumlah dokumen sebanyak 20 dengan isi setiap dokumen teks berbeda-beda dan dikelompokkan sebanyak 4 klaster. Berikut hasil pengujian:

**Tabel 5.7 Rekapitulasi Hasil Running Euclidian Distance 4**

No.	Euclidian Distance								Waktu (s)	Precision (%)
	Kelompok Yang Relevan				Kelompok Yang Tidak Relevan					
	1	2	3	4	1	2	3	4		
<b>I</b>	0	2	5	3	1	0	9	0	135.59	50
<b>II</b>	0	1	5	3	1	0	8	1	418.50	45
<b>III</b>	1	1	5	2	0	0	11	0	588.04	45
<b>IV</b>	5	1	2	1	11	0	0	0	315.20	45
<b>V</b>	2	3	3	2	6	0	4	0	319.24	50
<b>VI</b>	1	5	2	4	0	8	0	0	334.44	60
<b>VII</b>	1	1	2	5	0	0	3	8	353.87	45
<b>VIII</b>	2	2	2	3	0	0	4	7	139.75	45
<b>IX</b>	2	2	5	2	0	1	8	0	503.00	55
<b>X</b>	3	2	3	3	7	2	0	0	141.31	55
<b>Relevansi</b>										<b>49.5</b>



**5. Dokumen Teks 20 dengan 7 kelompok (Euclidian Distance)**

Pada pengujian ini diinputkan jumlah dokumen sebanyak 20 dengan isi setiap dokumen teks berbeda-beda dan dikelompokkan sebanyak 7 klaster. Berikut hasil pengujian:

**Tabel 5.8 Rekapitulasi Hasil Running Euclidian Distance 7**

No.	Euclidian Distance														Precision (%)	Keterangan	
	Kelompok Yang Relevan							Kelompok Yang Tidak relevan									
	1	2	3	4	5	6	7	1	2	3	4	5	6	7			
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
II	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
III	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VII	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VIII	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Pada hasil pengujian Euclidian Distance menggunakan 7 kelompok, proses iterasi k-means gagal. Hal ini disebabkan pada algoritma k-means sesuai langkah pembentukan klaster melalui tahapan iterasi, tahap iterasi dengan 7 kelompok proses iterasi tidak berhenti sehingga anggota dalam klaster tidak tetap, maka proses gagal di eksekusi.

**6. Dokumen teks 20 dengan 5 Kelompok (Euclidian Distance)**

Pada pengujian ini diinputkan jumlah dokumen sebanyak 20 dengan isi setiap dokumen teks telah diatur tingkat similarity nya mencapai 70%, dan dikelompokkan sebanyak 5 klaster. Berikut hasil pengujian :

**Tabel 5.9 Rekapitulasi Hasil Running Euclidian Distance 5**

No.	Euclidian Distance										Waktu (s)	Precision (%)
	Kelompok Yang Relevan					Kelompok Yang Tidak Relevan						
	1	2	3	4	5	1	2	3	4	5		
<b>I</b>	3	1	3	1	5	2	0	5	0	0	164.01	60
<b>II</b>	2	3	4	1	1	0	0	9	0	0	218.25	55
<b>III</b>	1	3	4	2	2	0	0	11	0	0	885.51	60
<b>Relevansi</b>												<b>58.33</b>

### 5.2.2 Hasil Pengujian

Dari hasil pengujian running di atas maka proses clustering sangat dipengaruhi oleh pembentukan centroid awal, sehingga proses dokumen cluster pada algoritma k-means selalu mengalami perubahan terhadap centroid yang random, sehingga dapat kita ambil nilai rata-rata dari nilai precision pengelompokan dengan menggunakan persamaan 2.11 pada bab 2. Dari hasil penghitungan nilai relevansi dapat kita buat tabel hasil pengujian relevansi seperti pada tabel berikut:

**Tabel 5.10 Hasil Pengujian Relevansi**

No.	Kelompok	Jaccard Coefficient (%)	Euclidian Distance (%)
<b>1</b>	4	73.5	49.5
<b>2</b>	7	79	Iterasi gagal
<b>3</b>	5	96.66	58.33

### 5.2.3 Pengujian *Blackbox*

Pengujian ini dilakukan untuk melihat apakah semua fungsi program telah berjalan sesuai dengan fungsi pengujian yang dihasilkan.

**Tabel 5.11 Pengujian *Blackbox* sistem *Jaccard Coefficient***

Deskripsi	Prosedur Pengujian	Masukan	Keluaran yang diharapkan	Kesimpulan
Pengujian Menu Home	Klik Menu Home	-	Menampilkan informasi mengenai text mining, algoritma winnowing, dan algoritma k-menas	Diterima
Pengujian Menu Corpus Dokumen	Klik Menu Corpus Dokumen	Menginputkan dokumen text yang akan diuji pada localhost c: dokumen teks	Menampilkan corpus seluruh dokumen teks yang telah di input	Diterima
Pengujian Menu Setting	Klik Menu Setting	<ol style="list-style-type: none"> <li>1. Input Kode uji</li> <li>2. Input nilai K-gram</li> <li>3. Input basis bilangan prima</li> <li>4. Input nilai window</li> <li>5. Input nilai kelompok</li> </ol>	Menampilkan inputan awal untuk proses dokumen clustering	Diterima
Pengujian Menu Clustering	Klik Menu Clustering	-	Menampilkan beberapa submenu: <ol style="list-style-type: none"> <li>1. Tampilkan proses</li> <li>2. Cek gram</li> <li>3. File cluster</li> </ol>	Diterima
Pengujian Menu List Uji	Klik Menu List Uji	-	Menampilkan hasil dari proses clustering seperti: <ol style="list-style-type: none"> <li>1. Kode uji</li> <li>2. K-gram</li> <li>3. Basis bilangan</li> <li>4. Window</li> <li>5. Kelompok</li> <li>6. Jumlah dokumen</li> <li>7. Waktu uji</li> </ol>	Diterima
Pengujian	Klik Menu	-	Menampilkan	Diterima

Menu About	About		informasi penulis dan cara penggunaan aplikasi dokumen clustering	
------------	-------	--	---	--

**Tabel 5.12 Pengujian *Blackbox* sistem *Euclidian Distance***

Deskripsi	Prosedur Pengujian	Masukan	Keluaran yang diharapkan	Kesimpulan
Pengujian Menu Home	Klik Menu Home	-	Menampilkan informasi mengenai text mining, algoritma winnowing, dan algoritma k-menas	Diterima
Pengujian Menu Corpus Dokumen	Klik Menu Corpus Dokumen	Menginputkan dokumen text yang akan diuji pada localhost c: dokumen teks	Menampilkan corpus seluruh dokumen teks yang telah di input	Diterima
Pengujian Menu Setting	Klik Menu Setting	<ol style="list-style-type: none"> <li>1. Input Kode uji</li> <li>2. Input nilai K-gram</li> <li>3. Input basis bilangan prima</li> <li>4. Input nilai window</li> <li>5. Input nilai kelompok</li> </ol>	Menampilkan inputan awal untuk proses dokumen clustering	Diterima
Pengujian Menu Clustering	Klik Menu Clustering	-	Menampilkan beberapa submenu: <ol style="list-style-type: none"> <li>1. Tampilkan proses</li> <li>2. Cek gram</li> <li>3. File cluster</li> </ol>	Diterima
Pengujian Menu List Uji	Klik Menu List Uji	-	Menampilkan hasil dari proses clustering seperti: <ol style="list-style-type: none"> <li>1. Kode uji</li> <li>2. K-gram</li> <li>3. Basis bilangan</li> <li>4. Window</li> <li>5. Kelompok</li> <li>6. Jumlah dokumen</li> <li>7. Waktu uji</li> </ol>	Diterima

Pengujian Menu About	Klik Menu About	-	Menampilkan informasi penulis dan cara penggunaan aplikasi dokumen clustering	Diterima
----------------------	-----------------	---	---	----------

#### 5.2.4 Kesimpulan Pengujian

1. Dari hasil pengujian konfigurasi pada tabel 5.3 hasil pengujian yang telah dilakukan pada algoritma winnowing diatas, untuk konfigurasi nilai yang digunakan adalah  $k=8$ ,  $b=3$ ,  $w=8$ , dimana proses komputasi untuk konfigurasi tersebut untuk algoritma winnowing telah memenuhi syarat untuk digunakan dalam metode pengelompokan k-means.
2. Dari hasil pengujian algoritma winnowing, seperti pada gambar 5.7 tampilan cek fingerprint relevan, dimana karakter winnowing adalah mencari fingerprint yang sama atau relevan antar dokumen, sehingga bila fingerprint yang didapat tidak muncul pada dokumen lain, maka memungkinkan kata yang muncul tidak identik dengan dokumen tersebut.
3. Dari hasil pengujian metode k-means pada tabel 5.4 sampai dengan tabel 5.9 rekapitulasi hasil running, dimana untuk menentukan centroid awal adalah acak (random), nilai awal centroid sangat berpengaruh terhadap pembentukan klaster, dan nilai anggota kelompok selalu berubah-ubah, sehingga diperlukan pengujian hasil running program beberapa kali.
4. Dari hasil running program pada tabel 5.10 hasil pengujian relevansi dapat kita ketahui nilai relevansi anggota kelompok yang akurat dan yang tidak akurat, dimana pada program untuk metode *jaccard coefficient* untuk pengujian I pada 4 kelompok nilai relevansi adalah 73,5 %, pada 7 kelompok 79 %, pada pengujian II pada 5 kelompok maka nilai relevansi 96,66 %.

Untuk metode *Euclidian distance* untuk pengujian I pada 4 kelompok nilai relevansi adalah 49,5 %, pada 7 kelompok nilai relevansi gagal, karena proses iterasi tidak berhenti, pada pengujian II pada 5 kelompok maka nilai relevansi 58,33 %.

5. Dari kesimpulan pengujian nomor 4 didapat metode *Jaccard coefficient* menghasilkan cluster yang lebih baik dari pada metode *Eclidian distance*, dimana dapat dilihat pada tabel 5.10 dan hasil relevansi yang lebih tinggi.
6. Pada pengujian *Blackbox* didapat fungsi dan tujuan telah sesuai dengan *output* sistem yang dihasilkan.