BAB 3
ANALISIS DAN PERANCANGAN

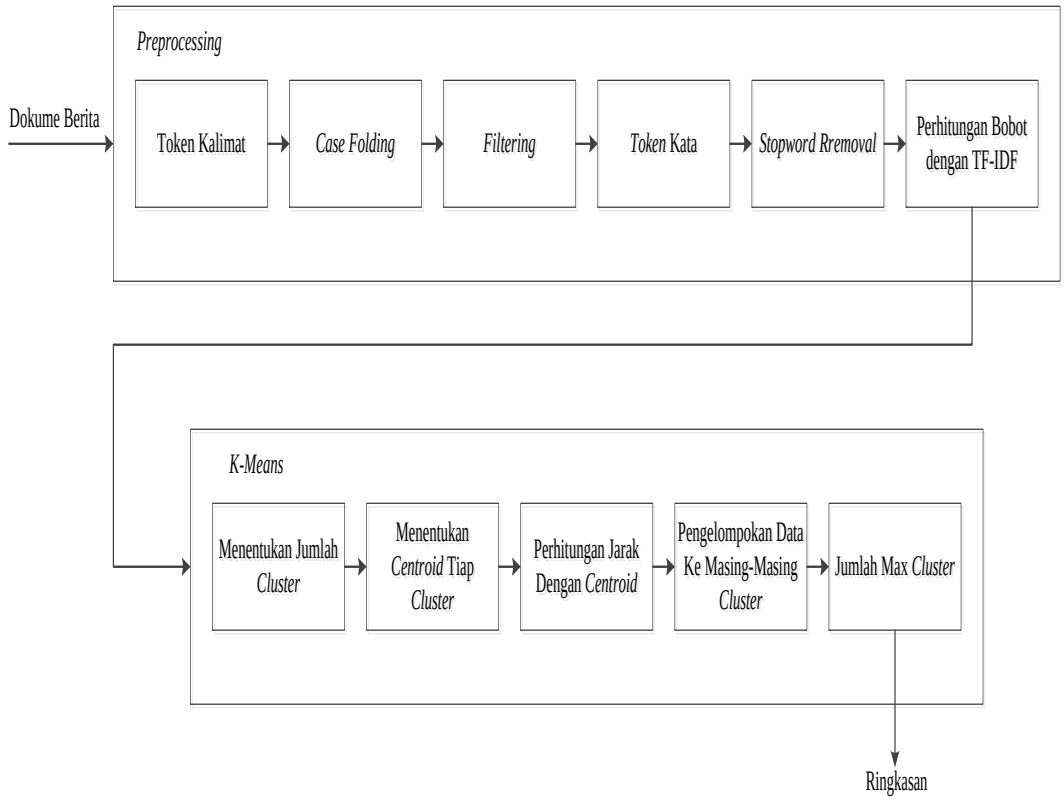
3.1 Analisis Masalah

Analisis masalah adalah gambaran masalah yang diangkat dalam penelitian tentang *text summarization* dalam artikel berita berbahasa Indonesia. Analisis masalah menjelaskan tentang proses identifikasi masalah serta evaluasi mengenai sistem peringkas teks otomatis dalam artikel berita berbahasa Indonesia.

Pada tugas akhir ini peringkasan teks otomatis(*automatic text summarization*) akan dilakukan dengan pendekatan statistik. Metode pertama yaitu *Term Frequency – Inverse Document Frequency* (TF-IDF) yang merupakan metode pemberian bobot hubungan suatu kata (*term*) terhadap dokumen. Bobot tiap kata (*term*) terhadap dokumen yang diperoleh akan dijumlahkan sehingga menjadi bobot kalimat[12]. Selanjutnya dari bobot tiap kalimat dijadikan data masukan proses *clustering* data dengan menggunakan metode *K-means Clustering*. Karena metode *K-means* dapat mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [8]. Pada tahap akhir, tiap bobot kalimat pada tiap *cluster* dijumlahkan selanjutnya dipilih jumlah yang paling besar. Nilai *cluster* paling besar itulah yang akan menjadi ringkasan dengan mengurutkan data aslinya, sehingga diharapkan hasil ringkasan yang diperoleh nantinya merupakan teks yang berisi informasi penting dari suatu bacaan dan panjang hasil ringkasan yang dihasilkan tersebut tidak lebihdari setengah panjang teks aslinya [13].

3.2 Arsitektur Sistem

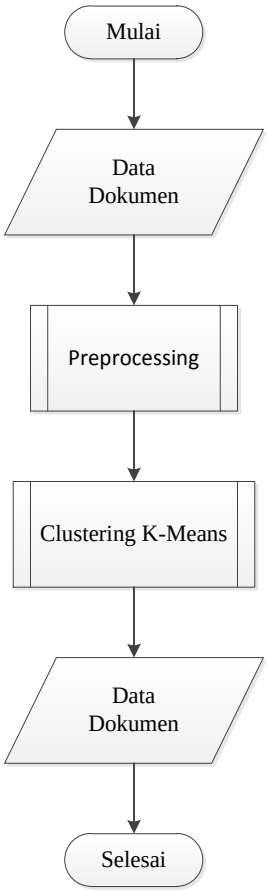
Arsitektur sistem menggambarkan perancangan sistem peringkasan teks otomatis secara keseluruhan. Adapun gambaran arsitektur sistem peringkasan teks otomasis seperti pada gambar 3.1 berikut.



Gambar 3. 1 Arsiteksur Sistem

3.3 Analisis Sistem

Analisis sistem didefinisikan sebagai penguraian dari sistem utama kedalam sub- sub sistem dengan tujuan untuk mendefinisikan permasalahan yang ada dan kebutuhan- kebutuhan yang diperlukan agar dapat diusulkan dan diciptakan sistem yang lebih baik. Implementasi metode *k-means* pada peringkasan teks yang akan dibangun bisa dilihat pada gambar 3.2 berikut.



Gambar 3. 2 Alur Sistem

Analisis sistem peringkasan teks otomatis yang akan dibangun pada penelitian ini memiliki 3 tahapan, yaitu *preprocessing*, penghitungan bobot dengan TF-IDF, *clustering K-Means*. Untuk penjelasan tiap tahapan dapat dilihat sebagai berikut :

1. Data dokumen

   Pada tahap ini pengguna memilih data yang akan diringkas. Data ini merupakan artikel berita berbahasa Indonesia dengan ektensi .txt.

2. *Preprocessing*

   Proses *preprocessing* merupakan tahap dimana data yang akan digunakan sebagai data testing dibersihkan dari hal yang tidak mempunyai pengaruh dalam ringkasan. *Preprocessing* yang akan dilakukan dalam tahap ini

meliputi : pemecahan kalimat, *case folding*(merubah menjadi *lowercase*), *filtering*, *tokenizing* dan hapus *stopword*. Kemudian dilanjutkan dengan pembobotan kata menggunakan TF-IDF.

3. *Clustering* dengan *K-means*

   Clustering yang kalimat yang berada pada satu kelas yang memiliki makna yang sama, sehingga nantinya akan dilakukan seleksi kalimat dengan cara mengambil salah satu kalimat dari setiap *cluster* sebagai perwakilan akan makna tertentu.

3.3.1 Analisis Data Masukan

Analisis data masukan *(input data)* pada peringkasan teks(*text summarization*) artikel berita berbahasa Indonesia menjelaskan proses data masukan yang dibuat merupakan sistem peringkasan dengan inputan data training berupa *single* dokumen untuk menghasilkan ringkasan (*summary*).

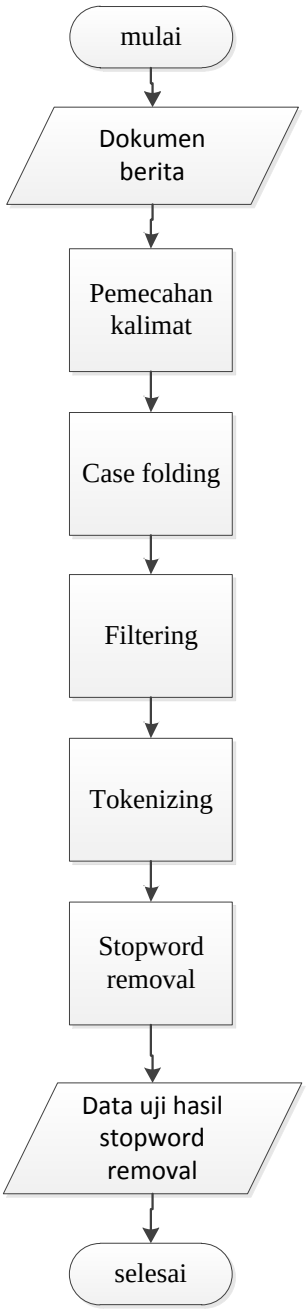
Data masukan pada penelitian ini didapat dari artikel berita megapolitan.kompas.com dengan judul "lulung dukung waarga bukit duri gugat pemprofDKI", *http://megapolotan.kompas.com/read/2016/05/13/10104761/Lulung.Dukung.Warga.Bukit.Duri.Gugat.Pemprov.DKI*

Tabel 3. 1 Data Masukan

Artikel Berita
Rencana warga Bukit Duri, Jakarta Selatan, menggugat Pemprov dan didukung Wakil Ketua DPRD Abraham "Lulung" Lunggana. Lulung menyatakan, gugatan memang harus dilakukan warga Bukit Duri untuk memberikan pelajaran bagi Pemprov DKI agar tidak sewenang-wenang. "Baguslah, Pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata Lulung di Masjid Luar Batang, Kamis (12/5/2016) malam. Ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan Monumen Nasional (Monas) dan kantor Balai Kota. Sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan Pasar Ikan yang telah digusur Pemprov DKI. "Dia (Ahok) jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya, RT, RW-nya, dan mereka bayar kewajiban pajak enggak? Ada nilai keekonomiannya, terus warga yang sudah bertahun-tahun di sana enggak boleh main gusur aja," ujar Lulung. Warga Bukit Duri memutuskan untuk mengajukan gugatan class action terhadap Pemerintah Provinsi DKI Jakarta yang berencana menertibkan permukiman tersebut. Gugatan telah didaftarkan ke Pengadilan Negeri Jakarta Pusat pada 10 Mei.

3.3.2 Analisis *Preprocessing*

Analisis *preprocessing* menjelaskan proses-proses yang dilakukan untuk mendapatkan fitur kata yang bersih sehingga tidak mengandung kata yang tidak pengaruh dalam ringkasan untuk menghasilkan proses selanjutnya. Berikut ini adalah tahapan *preprocessing* dalam sistem peringkasa teks artikel berita berbahasa Indonesia.



Gambar 3. 3 Tahap *Preprocessing*

Adapun penjelasan dari tahapan *preprocessing*. Berikut penjelasan dari tahapan- tahapan *preprocessing*:

1. Pemecahan Kalimat

Pada tahap pemecahan kalimat adalah memecah string menjadi kalimat-kalimat berdasarkan tanda titik (.) sebagai pemisah. Hasil pemecahan dokumen menjadi kalimat-kalimat dapat dilihat pada Tabel 3.2 berikut.

Tabel 3. 2 Pemecahan Kalimat

No	Kalimat
S1	Rencana warga Bukit Duri, Jakarta Selatan, menggugat Pemprov dan didukung Wakil Ketua DPRD Abraham "Lulung" Lunggana
S2	Lulung menyatakan, gugatan memang harus dilakukan warga Bukit Duri untuk memberikan pelajaran bagi Pemprov DKI agar tidak sewenang-wenang
S3	"Baguslah, Pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata Lulung di Masjid Luar Batang, Kamis (12/5/2016) malam.
S4	Ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan Monumen Nasional (Monas) dan kantor Balai Kota.
S5	Sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan Pasar Ikan yang telah digusur Pemprov DKI.
S6	. "Dia (Ahok) jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya, RT, RW-nya, dan mereka bayar kewajiban pajak enggak? Ada nilai keekonomiannya, terus warga yang sudah bertahun-tahun di sana enggak boleh main gusur aja," ujar Lulung.
S7	Warga Bukit Duri memutuskan untuk mengajukan gugatan class action terhadap Pemerintah Provinsi DKI Jakarta yang berencana menertibkan permukiman tersebut.
S8	Gugatan telah didaftarkan ke Pengadilan Negeri Jakarta Pusat pada 10 Mei.

Keterangan

S1 = Kalimat ke-1.

S2 = Kalimat ke-2.

S3 = Kalimat ke-3.

S4 = Kalimat ke-4.

S5 = Kalimat ke-5.

S6 = Kalimat ke-6.

S7 = Kalimat ke-7

S8 = Kalimat ke-8.

2. *Case Folding*

Pada tahap *casefolding* adalah membuat karakter/huruf menjadi *lowercase*. Hasil dari *case folding* dapat dilihat pada Tabel 3.3 berikut.

Tabel 3. 3 *Case Folding*

No	Kalimat	Hasil *case folding*
S1	Rencana warga Bukit Duri, Jakarta Selatan, menggugat Pemprov dan didukung Wakil Ketua DPRD Abraham "Lulung" Lunggana	rencana warga bukit duri, jakarta selatan, menggugat pemprov dan didukung wakil ketua dprd abraham "lulung" lunggana
S2	Lulung menyatakan, gugatan memang harus dilakukan warga Bukit Duri untuk memberikan pelajaran bagi Pemprov DKI agar tidak sewenang-wenang	lulung menyatakan, gugatan memang harus dilakukan warga bukit duri untuk memberikan pelajaran bagi pemprov dki

		agar tidak sewenang-wenang
S3	"Baguslah, Pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata Lulung di Masjid Luar Batang, Kamis (12/5/2016) malam.	"baguslah, pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata lulung di masjid luar batang, kamis (12/5/2016) malam.
S4	Ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan Monumen Nasional (Monas) dan kantor Balai Kota.	ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan monumen nasional (monas) dan kantor balai kota.
S5	Sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan Pasar Ikan yang telah digusur Pemprov DKI.	sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan pasar ikan yang telah digusur pemprov dki.
S6	"Dia (Ahok) jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya, RT, RW-nya, dan mereka bayar kewajiban pajak enggak? Ada nilai keekonomiannya, terus warga yang sudah bertahun-tahun di sana enggak boleh main gusur aja," ujar Lulung.	"dia (ahok) jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya, rt, rw-nya, dan mereka bayar kewajiban pajak enggak? ada nilai keekonomiannya, terus warga yang sudah bertahun-tahun di sana enggak boleh main gusur aja," ujar lulung.
S7	Warga Bukit Duri memutuskan untuk mengajukan gugatan class action terhadap	warga bukit duri memutuskan untuk mengajukan gugatan

	Pemerintah Provinsi DKI Jakarta yang berencana menertibkan permukiman tersebut.	class action terhadap pemerintah provinsi dki jakarta yang berencana menertibkan permukiman tersebut.
S8	Gugatan telah didaftarkan ke Pengadilan Negeri Jakarta Pusat pada 10 Mei.	gugatan telah didaftarkan ke pengadilan negeri jakarta pusat pada 10 mei.

Keterangan

S1 = Kalimat ke-1.

S2 = Kalimat ke-2.

S3 = Kalimat ke-3.

S4 = Kalimat ke-4.

S5 = Kalimat ke-5.

S6 = Kalimat ke-6.

S7 = Kalimat ke-7

S8 = Kalimat ke-8.

3. Filtering

Filtering digunakan untuk menghilangkan karakter – karakter non – huruf yang dianggap tidak valid (karakter angka dan simbol). Karakter – karakter yang akan dihapus atau dianggap sebagai pemisah kata, dapat dilihat pada Tabel 3.4 berikut.

Tabel 3. 4 Karakter yang akan dihapus

Karakter			
!	(	<	]
@	)	:	|
#	.	;	\
$	,	/	+
%	‘	?	=
^	“	{	_
&	-	}	,
*	>	[	Tab, spasi, enter

Hasil filtering dapat dilihat pada Tabel 3.5 berikut.

Tabel 3. 5 Hasil *Filtering*

No	Data hasil *case folding*	Seteleh *Filtering*
S1	rencana warga bukit duri, jakarta selatan, menggugat pemprov dan didukung wakil ketua dprd abraham "lulung" lunggana	rencana warga bukit duri jakarta selatan menggugat pemprov dan didukung wakil ketua dprd abraham lulung lunggana
S2	lulung menyatakan, gugatan memang harus dilakukan warga bukit duri untuk memberikan pelajaran bagi pemprov dki agar tidak sewenang-wenang	lulung menyatakan gugatan memang harus dilakukan warga bukit duri untuk memberikan pelajaran bagi pemprov dki agar tidak sewenang wenang
S3	"baguslah, pemprov harus mengerti mana yang menjadi tanah negara,	baguslah pemprov harus mengerti mana yang menjadi tanah negara

	mana yang dikelola oleh rakyat," kata lulung di masjid luar batang, kamis (12/5/2016) malam.	mana yang dikelola oleh rakyat kata lulung di masjid luar batang kamis malam
S4	ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan monumen nasional (monas) dan kantor balai kota.	ia mencontohkan tanah negara yang menjadi aset pemerintah seperti kawasan monumen nasional monas dan kantor balai kota
S5	sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan pasar ikan yang telah digusur pemprov dki.	sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan pasar ikan yang telah digusur pemprov dki
S6	"dia (ahok) jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya, rt, rw-nya, dan mereka bayar kewajiban pajak enggak? ada nilai keekonomiannya, terus warga yang sudah bertahun-tahun di sana enggak boleh main gusur aja," ujar lulung.	dia ahok jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya rt rw nya dan mereka bayar kewajiban pajak enggak ada nilai keekonomiannya terus warga yang sudah bertahun tahun di sana enggak boleh main gusur aja ujar lulung
S7	warga bukit duri memutuskan untuk mengajukan gugatan class action terhadap pemerintah provinsi dki jakarta yang berencana menertibkan permukiman tersebut.	warga bukit duri memutuskan untuk mengajukan gugatan class action terhadap pemerintah provinsi dki jakarta yang berencana menertibkan permukiman tersebut
S8	gugatan telah didaftarkan ke pengadilan negeri jakarta pusat pada 10 mei.	gugatan telah didaftarkan ke pengadilan negeri jakarta pusat pada 10 mei

Keterangan

S1 = Kalimat ke-1.

S2 = Kalimat ke-2.

S3 = Kalimat ke-3.

S4 = Kalimat ke-4.

S5 = Kalimat ke-5.

S6 = Kalimat ke-6.

S7 = Kalimat ke-7

4. Tokenizing

Pada tahap *Tokenizing* kata adalah menghilangkan karakter pemisah yang menyusunya berupa katakter spasi. Hasil *tokenizing* dapat dilihat pada Tabel 3.6 berikut..

Tabel 3. 6 *Tokenizing*

No	Data hasil *case folding*	*Tokenizing*	
S1	rencana warga bukit duri jakarta selatan menggugat pemprov dan didukung wakil ketua dprd abraham lulung lunggana	rencana	didukung
		warga	wakil
		bukit	ketua
		duri	dprd
		jakarta	abraham
		selatan	lulung
		menggugat	Lunggana
		pemprov	
		dan	

Keterangan

S1 = Kalimat ke-1

(Selengkapnya Lihat Lampiran A.)

5. Stopword Removal

Proses *stopword removal* adalah proses menghapus kata yang tidak penting yang diambil dari tahap *tokenizing*. Pada tahap *stopword* ini menggunakan algoritma *stoplist* dengan membuang kata yang kurang penting. Contoh *stopword* adalah"di", "dari", "yang" dan lain-lain. Hasil dari *stop word removal* dapat dilihat pada Tabel 3.8. Adapun isi dari sebagian daftar kata stopword yang disajikan pada Tabel 3.7 berikut.

Tabel 3. 7 Daftar Sebagian *Stopword Removal*

telah	punya	mendapatkan	Dari
antara	agar	juga	Semua
setelah	jadi	ada	Setelah

(Selengkapnya Lihat Lampiran B)

Tabel 3. 8 Hasil *Stopword Removal*

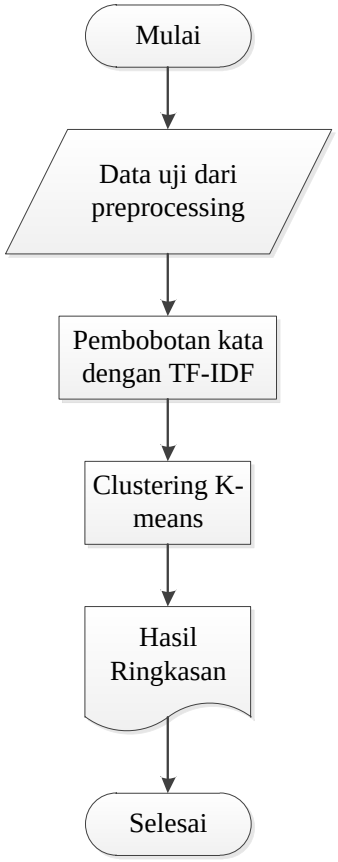
No	Data hasil *tokenizing*		*stopword*	
S1	rencana	didukung	Rencana	didukung
	warga	wakil	warga	wakil
	bukit	ketua	bukit	ketua
	duri	dprd	duri	dprd
	jakarta	abraham	jakarta	abraham
	selatan	lulung	selatan	lulung

	menggugat	lunggana	menggugat	lunggana
	pemprov		pemprov	
	dan			

(Selangkapnya Lihat Lampiran C)

3.3.3 Analisis *Summarization*

Pada alanisis *simmarization* akan menjelaskan proses peringkasan teks artikel berita berbahasa Indonesia dengan metode pembobotan kata TF-IDF dan metode *Clustering K-Means*. Berikut adalah *flowchart* proses peringkasan teks berita artikel berbahasa Indonesia dengan metode TF-IDF dan metode *Clustering K-Means*:



Gambar 3. 4 *Flowchart* peringkasan dengan metode TF-IDF dan *K-Means*

3.3.4 Algoritma TD-IDF

Hasil proses text *preprosessing* dilakukan pembobotan TF-IDF. Pembobotan secara otomatis biasanya berdasarkan jumlah kemunculan suatu kata dalam sebuah dokumen (*term frequency*) dan jumlah kemunculannya dalam koleksi dokumen (*inverse document frequency*). Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen. Rumus TF-IDF dapat dilihat pada persamaan 2.2 Proses awal dilakukan perhitungan kata (*term*) pada tiap dokumen, sehingga mendapatkan *frekuensi term* (*tf*). Selanjutnya mencari *documen frequency (df), df* merupakan banyaknya dokumen dimana satu kata (*term*) muncul. Contoh kata "warga" maka.

Diketahui

Jumlah kalimat ($S_{(n)}$) = 8,

df = 4 (karena kata "warga" muncul di S1,S2,S6,S7),

Idf(t) = log(D/*df*) = log(8/4) = 0.301.

Untuk lebih jelasnya dapat dilihat pada Tabel 3.9 berikut.

Tabel 3. 9 Perhitungan IDF Dari Tiap Kata(*Term*).

Term	*Tf*								*df*	$S_{(n)}/df$	Idf =log($S_{(n)}/df$)
	S1	S2	S3	S4	S5	S6	S7	S8			
rencana	1	0	0	0	0	0	0	0	1	8	0.903
Warga	1	1	0	0	0	1	1	0	4	2	0.301
Bukit	1	1	0	0	0	0	1	0	3	2.67	0.426
Duri	1	1	0	0	0	0	1	0	3	2.67	0.426
...	...	...	...	...	...	...	...	...	...	...	...
Pusat	0	0	0	0	0	0	0	1	1	8	0.903
mei	0	0	0	0	0	0	0	1	1	8	0.903

Keterangan

Term = Daftar kata yang terdapat pada artikel.

S1 = Kalimat ke-1.

S2 = Kalimat ke-2.

S3 = Kalimat ke-3.

S4 = Kalimat ke-4.

S5 = Kalimat ke-5.

S6 = Kalimat ke-6.

S7 = Kalimat ke-7

S8 = Kalimat ke-8.

df = Frekuensi Dokumen *(dokumen frequence).*

$S_{(n)}$ = Jumlah *Sentence/*Kalimat.

Idf = Perhitungan jumlah dokumen yang mengandung sebuah kata (*term*) yang dicari dari kumpulan dokumen yang ada.

(Selengkapnya Lihat Lampiran D)

Untuk memperoleh bobot pada kata "warga" maka dilakukan perhitungan perkalian *tf* dengan *idf*

$tf\text{-}idf_{(t)} = tf_{(t,d)}$ x $idf_{(t)}$ = 1 x 0.301 = 0.301

untuk lebih jelasnya dapat dilihat pada tabel 3.6 berikut.

Tabel 3. 10 Perhitungan TF-IDF

Term	Kalimat (S)							
	S1	S2	S3	S4	S5	S6	S7	S8
Rencana	0.903	0	0	0	0	0	0	0
Warga	0.301	0.301	0	0	0	0.301	0.301	0
Bukit	0.426	0.426	0	0	0	0	0.426	0.477
Duri	0.426	0.426	0	0	0	0	0.426	0.477
Jakarta	0.426	0	0	0	0	0	0.426	0.426
Selatan	0.903	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
Pusat	0	0	0	0	0	0	0	0.903
Mei	0	0	0	0	0	0	0	0.903
$TF\text{-}IDF_{(w)}$	10.91	5.316	9.155	9.145	8.678	15.653	10.257	5.969

Keterangan

Term = Daftar kata yang terdapat pada artikel.

S1 = Kalimat ke-1.

S2 = Kalimat ke-2.

S3 = Kalimat ke-3.

S4 = Kalimat ke-4.

S5 = Kalimat ke-5.

S6 = Kalimat ke-6.

S7 = Kalimat ke-7

S8 = Kalimat ke-8.

$TF\text{-}IDF_{(w)}$ = Bobot kalimat.

(Selengkapnya Lihat Lampiran E)

3.3.5 Algoritma *K-Means*

Dari hasil pembobotan kalimat dengan TF-IDF pada topik artikel tersebut selanjutnya dijumlahkan nilai tiap kalimat yang digunakan sebagai data inputan algoritma *K-Means Clustering* [13]. untuk lebih jelasnya perhatikan tabel 3.11 berikut.

Tabel 3. 11 Data Inputan *K-Means*

Kalimat (S)	**$TF\text{-}IDF_{(w)}$**
S1	10.91
S2	5.316
S3	9.155
S4	9.145
S5	8.678
S6	15.653
S7	10.257
S8	5.969

Keterangan

S1 = Kalimat ke-1.

S2 = Kalimat ke-2.

S3 = Kalimat ke-3.

S4 = Kalimat ke-4.

S5 = Kalimat ke-5.

S6 = Kalimat ke-6.

S7 = Kalimat ke-7

S8 = Kalimat ke-8.

TF-IDF = hasil perhitungan TF-IDF pada tiap kalimat.

Penentuan Jumlah *Cluster*

Adapun cara untuk menentukan jumlah *cluster* (K) yaitu dengan membagi 2 jumlah kalimat yang ada pada berita artikel kemudian hasil dari pembagian diakarkuadrat[12]. Pada kasus ini terdapat 8 jumlah kalimat berikut untuk lebih jelasnya.

$$K = \sqrt{S(n) / 2} \quad (3.1)$$

Maka diperoleh

$K = \sqrt{8 / 2}$ = 2

Dimana :

K = jumlah cluster/kelompok

$S(n)$ = Jumlah Kalimat

Jika hasil dari pembagian berupa bilangan desimal maka dibulatkan bawah. Misalnya terdapat 10 kalimat maka

$K = \sqrt{10 / 2}$ = 2,23 dibulatkan kebawah menjadi 2

Menentukan Nilai *Centroid* / titik pusat cluster

Dalam menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Contohnya pada kasus ini terpilih bobot kalimat(TF-IDF) ke-1 dan ke-5. Misalnya :

Diambil data **ke-1** sebagai pusat ***cluster* ke-1 (wC1)** = 10.91

Diambil data **ke-5** sebagai pusat ***cluster* ke-2 (wC2)** = 8.678

Keterangan

wC1 = Cluster (kelompok) ke-1

wC2 = Cluster (kelompok) ke-2

Menghitung Jarak Antara Data Dengan Pusat C*luster*.

Untuk menghitung jarak dengan pusat *Cluster* tersebut dapat menggunakan persamaan (2.3) agar nilai yang dihasilkan selalu positif. *Centroid* terdekat akan menjadi cluster yang diikuti oleh data tersebut. Kemudian akan didapatkan matrik jarak yaitu wC1 dan wC2. Berikut perhitungan jarak antara data ke setiap centroid.

Perhitungan pada Cluster 1 (wC1)

D*e* (w1,wC1) = $\sqrt{(10.91 - 10.91)^2}$ = 0

D*e* (w2,wC1) = $\sqrt{(5.316 - 10.91)^2}$ = 5.594

D*e* (w3,wC1) = $\sqrt{(9.155 - 10.91)^2}$ = 1.755

D*e* (w4,wC1) = $\sqrt{(9.145 - 10.91)^2}$ = 1.454

D*e* (w5,wC1) = $\sqrt{(8.678 - 10.91)^2}$ = 2.232

D*e* (w6,wC1) = $\sqrt{(15.653 - 10.91)^2}$ = 4.473

D*e* (w7,wC1) = $\sqrt{(10.257 - 10.91)^2}$ = 0.653

D*e* (w8,wC1) = $\sqrt{(5.969 - 10.91)^2}$ = 4.941

Perhitungan pada Cluster 2 (wC2)

D*e* (w1,wwC2) = $\sqrt{(10.91 - 8.678)^2}$ = 2.232

D*e* (w2,wC2) = $\sqrt{(5.316 - 8.678)^2}$ = 3.362

D*e* (w3,wC2) = $\sqrt{(9.155 - 8.678)^2}$ = 0.477

D*e* (w4,wC2) = $\sqrt{(9.145 - 8.678)^2}$ = 0.778

D*e* (w5,wC2) = $\sqrt{(8.678 - 8.678)^2}$ = 0

D*e* (w6,wC2) = $\sqrt{(15.653 - 8.678)^2}$ = 6.975

D*e* (w7,wC2) = $\sqrt{(10.257 - 8.678)^2}$ = 1.579

D*e* (w8,wC2) = $\sqrt{(5.969 - 8.678)^2}$ = 2.709

Setelah mendapatkan nilai dari masing-masing cluster, kemudian ditetapkan jarak terpendeknya, untuk mengetahui bobot kalimat(TF-IDF) tersebut berada di salah satu cluster(kelompok). Untuk lebih jelasnya perhatikan tabel 3.12 berikut.

Tabel 3. 12 ***Euclidian Distance Iterasi 1***

No	S (Kalimat)	$TF\text{-}IDF_{(w)}$	wC1	wC2	Jarak Terpendek	Keterangan
1	**S1**	10.91	0	2.232	0	wC1
2	**S2**	5.316	5.594	3.362	3.362	wC2
3	**S3**	9.155	1.755	0.477	0.477	wC2

4	**S4**	9.145	1.454	0.778	0.778	wC2
5	**S5**	8.678	2.232	0	0	wC2
6	**S6**	15.653	4.473	6.975	4.473	wC1
7	**S7**	10.257	0.653	1.579	0.653	wC1
8	**S8**	5.969	4.941	2.709	2.709	wC2

Pengelompokan Data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Berikut data matrik penglompokan grup, nilai 1 berati data tersebut berada dalam grup tersebut. Untuk lebih jelasnya perhatikan tabel 3.13 berikut.

Tabel 3. 13 Pengelompokan Data Iterasi ke-1

S (Kalimat)	wC1	wC2
S1	1	
S2		1
S3		1
S4		1
S5		1
S6	1	
S7	1	
S8		1

Penentuan Pusat *Cluster* Baru

Setelah anggota tiap *cluster* diketahui kemudian menentukan pusat *cluster* baru berdasarkan nilai rata-rata dari data yang terletak pada *centroid* yang sama. Perhitungan pusan *cluster* baru dapat dilihat pada tabel 3.14 berikut.

Tabel 3. 14 Pusat Cluster Baru Iterasi ke-1

S / Kalimat	TF-IDF$_{(w)}$	wC1	wC2
S1	10.91	10.91	
S2	5.316		5.316
S3	9.155		9.155
S4	9.145		9.145
S5	8.678		8.678
S6	15.653	15.653	
S7	10.257	10.257	
S8	5.969		5.969
Centroid Baru		12.273	7.714

Iterasi Ke-2

Untuk iterasi ke-2 maka proses dilakukan dengan menentukan titik pusat cluster, kemudian menghitung jarak antara data dengan pusat c*luster* hingga posisi data tidak mengalami perubahan.

Centroid baru w**C1** = 12.273

Centroid baru w**C2** = 7.714

Menghitung Jarak Antara Data Dengan Pusat C*luster* Pada Iterasi ke-2

Proses penghitungan jarak antara data dengan pusat cluster dapat dilihat pada proses berikut ini.

Perhitungan pada *Cluster* 1 (wC1)

D*e* (w1,wC1) = $\sqrt{(10.91 - 12.273)^2}$ = 1.363

D*e* (w2,wC1) = $\sqrt{(5.316 - 12.273)^2}$ = 6.957

D*e* (w3,wC1) = $\sqrt{(9.155 - 12.273)^2}$ = 3.118

D*e* (w4,wC1) = $\sqrt{(9.145 - 12.273)^2}$ = 2.817

D*e* (w5,wC1) = $\sqrt{(8.678 - 12.273)^2}$ = 3.595

D*e* (w6,wC1) = $\sqrt{(15.653 - 12.273)^2}$ = 3.379

D*e* (w7,wC1) = $\sqrt{(10.257 - 12.273)^2}$ = 2.016

D*e* (w8,wC1) = $\sqrt{(5.969 - 12.273)^2}$ = 6.304

Perhitungan pada Cluster 2 (wC2)

D*e* (w1,wC2) = $\sqrt{(10.91 - 7.714)^2}$ = 3.195

D*e* (w2,wC2) = $\sqrt{(5.316 - 7.714)^2}$ = 2.399

D*e* (w3,wC2) = $\sqrt{(9.155 - 7.714)^2}$ = 1.44

D*e* (w4,wC2) = $\sqrt{(9.145 - 7.714)^2}$ = 1.741

D*e* (w5,wC2) = $\sqrt{(8.678 - 7.714)^2}$ = 0.963

D*e* (w6,wC2) = $\sqrt{(15.653 - 7.714)^2}$ = 7.938

D*e* (w7,wC2) = $\sqrt{(10.257 - 7.714)^2}$ = 2.542

D*e* (w8,wC2) = $\sqrt{(5.969 - 7.714)^2}$ = 1.746

Setelah mendapatkan nilai dari masing-masing cluster, kemudian ditetapkan jarak terpendeknya. Untuk lebih jelasnya perhatikan tabel 3.15 berikut.

Tabel 3. 15 *Euclidian Distance* Iterasi ke-2

No	S (Kalimat)	TF-IDF$_{(w)}$	wC1	wC2	Jarak Terpendek	Keterangan
1	**S1**	10.91	1.363	3.195	1.363	wC1
2	**S2**	5.316	6.957	2.399	2.399	wC2
3	**S3**	9.155	3.118	1.44	1.44	wC2
4	**S4**	9.145	2.817	1.741	1.741	wC2
5	**S5**	8.678	3.595	0.963	0.963	wC2
6	**S6**	15.653	3.379	7.938	3.379	wC1
7	**S7**	10.257	2.016	2.542	2.016	wC1
8	**S8**	5.969	6.304	1.746	1.746	wC2

Pengelompokan Data

Berikut data matrik penglompokan data, nilai 1 berati data tersebut berada dalam grup tersebut. Kemudian bandingkan pengelompokan data ke-1 dan ke-2, jika posisi anggota *cluster* tetap atau tidak berubah maka proses perhitungan selesai, sebaliknya jika posisi anggota *cluster* berubah maka ditetapkan pusat *cluster* baru dan dilakukan perhitungan antara data denga pusat *cluster* baru. Untuk lebih jelasnya perhatikan tabel 3.16 berikut

Tabel 3. 16 perbandingan Data Iterasi ke-1 dan Data Iterasi ke-2

Data Iterasi ke-1			**Data Iterasi ke-2**		
S (Kalimat)	**wC1**	**wC2**	**S (Kalimat)**	**wC1**	**wC2**
S1	1		**S1**	1	
S2		1	**S2**		1
S3		1	**S3**		1
S4		1	**S4**		1
S5		1	**S5**		1
S6	1		**S6**	1	
S7	1		**S7**	1	
S8		1	**S8**		1

Karena hasil Clustering pengelompokan data ke-1 sama dengan pengelompokan data ke-2 maka tidak perlu dilakukan perhitungan lagi. Maka selanjutnya mendapatkan hasil ringkasan, yaitu dengan menjumlahakan data pada masing-masing *cluster* nilai yang tertinggi yang akan menjadi hasil ringkasan berdasarkan urutan data asli [13].

Tabel 3. 17 Hasil Penjumlahan Pada Tiap-Tiap Cluster

No	**S (Kalimat)**	**TF-IDF$_{(w)}$**	**wC1**	**wC2**
1	**S1**	10.91	10.91	
2	**S2**	5.316		5.316
3	**S3**	9.155		9.155
4	**S4**	9.145		9.145
5	**S5**	8.678		8.678
6	**S6**	15.653	15.653	
7	**S7**	10.257	10.257	
8	**S8**	5.969		5.969
Total			36.82	38.574

Hasil dari penjumlahan masing-masing cluster adalah wC1 memiliki jumlah 36.82 dan wC2 memiliki jumlah 38.574, sehingga wC2 menunjukkan nilai tertinggi. Maka untuk mendapatkan hasil ringkasan berada pada cluster wC2 dengan anggota S2,S3,S4,S5,dan S8. Untuk lebih jelasnya dapat dilihat pada tabel berikut.

Tabel 3. 18 Data *Cluster* C2

No	Kalimat
S2	Lulung menyatakan, gugatan memang harus dilakukan warga Bukit Duri untuk memberikan pelajaran bagi Pemprov DKI agar tidak sewenang-wenang
S3	"Baguslah, Pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata Lulung di Masjid Luar Batang, Kamis (12/5/2016) malam.
S4	Ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan Monumen Nasional (Monas) dan kantor Balai Kota.
S5	Sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan Pasar Ikan yang telah digusur Pemprov DKI.
S8	Gugatan telah didaftarkan ke Pengadilan Negeri Jakarta Pusat pada 10 Mei.

Untuk hasil ringkasan dapat dilihat pada tabel berikut ini**.**

Tabel 3. 19 Hasil Ringkasan

Dokumen asli	Hasil ringkasan
Rencana warga Bukit Duri, Jakarta Selatan, menggugat Pemprov dan didukung Wakil Ketua DPRD Abraham "Lulung" Lunggana. Lulung	Lulung menyatakan, gugatan memang harus dilakukan warga Bukit Duri untuk memberikan pelajaran bagi Pemprov DKI agar tidak sewenang-

menyatakan, gugatan memang harus dilakukan warga Bukit Duri untuk memberikan pelajaran bagi Pemprov DKI agar tidak sewenang-wenang. "Baguslah, Pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata Lulung di Masjid Luar Batang, Kamis (12/5/2016) malam. Ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan Monumen Nasional (Monas) dan kantor Balai Kota. Sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan Pasar Ikan yang telah digusur Pemprov DKI. "Dia (Ahok) jangan mengklaim saja, lihat dulu di sana ada enggak penduduknya, RT, RW-nya, dan mereka bayar kewajiban pajak enggak? Ada nilai keekonomiannya, terus warga yang sudah bertahun-tahun di sana enggak boleh main gusur aja," ujar Lulung. Warga Bukit Duri memutuskan untuk mengajukan gugatan class action terhadap Pemerintah Provinsi DKI Jakarta yang berencana menertibkan permukiman tersebut. Gugatan telah	wenang. "Baguslah, Pemprov harus mengerti mana yang menjadi tanah negara, mana yang dikelola oleh rakyat," kata Lulung di Masjid Luar Batang, Kamis (12/5/2016) malam. Ia mencontohkan tanah negara yang menjadi aset pemerintah, seperti kawasan Monumen Nasional (Monas) dan kantor Balai Kota. Sementara lahan negara yang dikelola oleh masyarakat salah satunya di kawasan Pasar Ikan yang telah digusur Pemprov DKI. Gugatan telah didaftarkan ke Pengadilan Negeri Jakarta Pusat pada 10 Mei.

didaftarkan ke Pengadilan Negeri Jakarta Pusat pada 10 Mei.	

3.4 Analisis Kebutuhan Non-Fungsional

Analisis kebutuhan non-fungsional pada penelitian ini terdiri dari kebutuhan perangkat keras, kebutuhan perangkat lunak, dan kebutuhan pengguna.

3.4.1 Kebutuhan Perangkat Keras *(hardware)*.

Kebutuhan perangkat keras yang digunakan dalam pembangunan *text summarization* menggunakan metode *K-Means* pada artikel berita berbahasa Indonesia adalah dapat dilihat pada tabel 3.20 berikut.

Tabel 3. 20 Kebutuhan Perangkat Keras

Perangkat Keras	Spesifikasi
Processor	Intel(R) Dual Core 1.5GHz
RAM	2048 MB
Monitor	10.1 inch
Keyboard	Standar
Mouse	Standar

3.4.2 Kebutuhan Perangkat Lunak (*Software*)

kebutuhan perangkat lunak (*software*) yang digunakan dalam pembangunan text summarization dengan metode K-Means pada artikel berita berbahasa Indonesia adalah sebagai berikut.

1. Sistem Operasi Windows Ultimate 7 32bit
2. Web Browser Google Chrome
3. Code editor SublimeText3 Portable
4. Bahasa Pemrograman PHP
5. XAMPP.

3.4.3 Kebutuhan Pengguna

Kebutuhan pengguna pada *text summarization* menggunakan metode MMR pada artikel berbahasa Indonesia adalah sebagai berikut:

a. Menguasai penggunaan komputer.
b. Mengerti secara teknis *tools* dan *software* pendukung dalam menjalankan aplikasi.

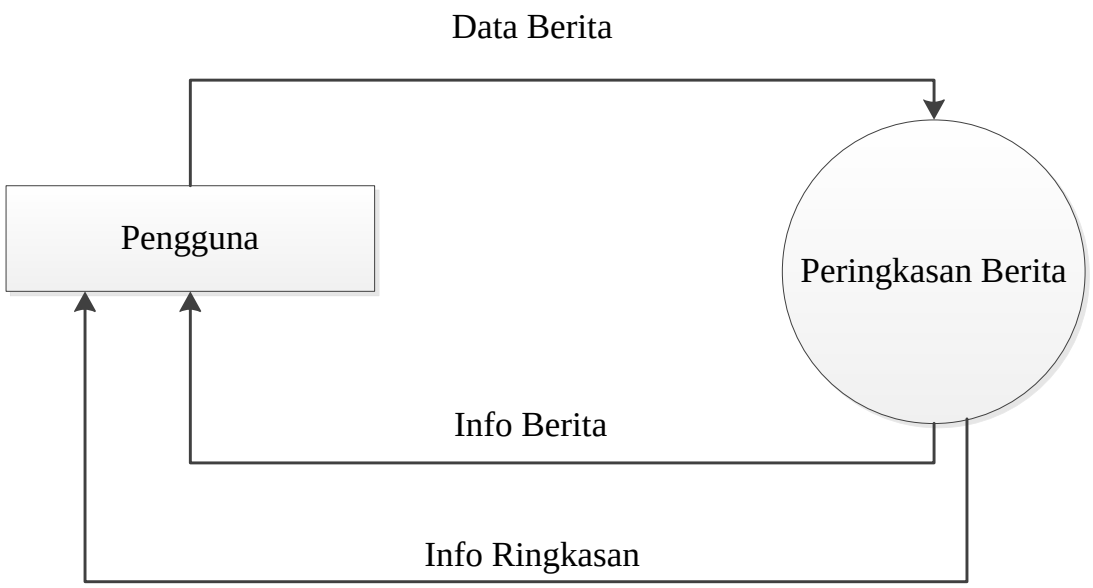
   Mengerti tahap-tahap dalam menjalankan aplikasi *text summarization.*

3.5 Analsisi Kebutuhan Fungsional

Analisis kebutuhan fungsional bertujuan untuk menganalisis proses yang akan diterapkan dalam sistem yang akan dibangun. Analisis kebutuhan fungsional pada aplikasi ini menggunakan Diagram konteks,pembuatan DFD, Spesifikasi proses, dan kamus data.

3.5.1 Diagram Konteks

Diagram konteks atau disebut juga dengan model sistem fundamental mempresentasikan seluruh elemen sistem sebagai sebuah bubble tunggal dengan data input output yang di tunjukan oleh arah panah yang masuk dan keluar secara berurutan. Berdasarkan hasil analisis maka sistem yang dibangun dapat di lihat dalam diagram konteks berikut:



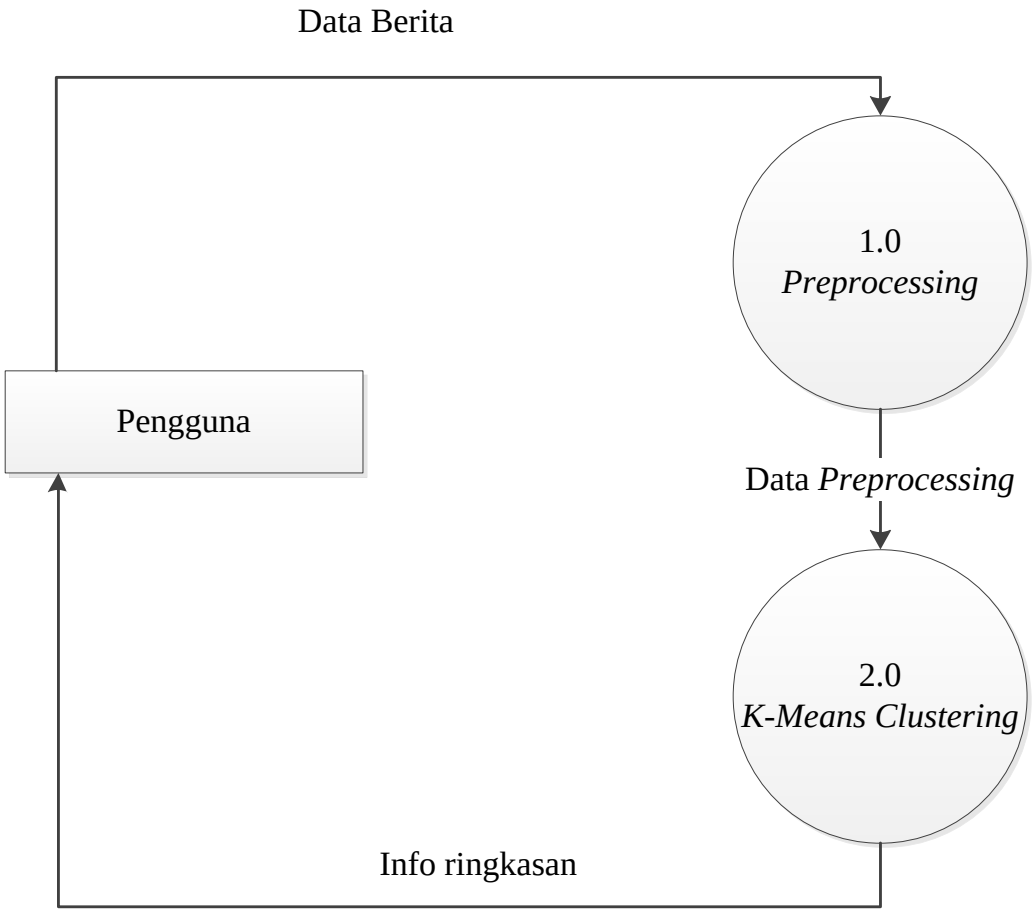
Gambar 3. 5 Diagram konteks

3.5.2 Data Flow Diagram (DFD)

Data Flow Diagram (DFD) merupakan diagram konteks dalam bentuk yang lebih detail. DFD menguraikan proses-proses yang terjadi dalam sistem sampai prosesyang lebih detail.

1. DFD level 1

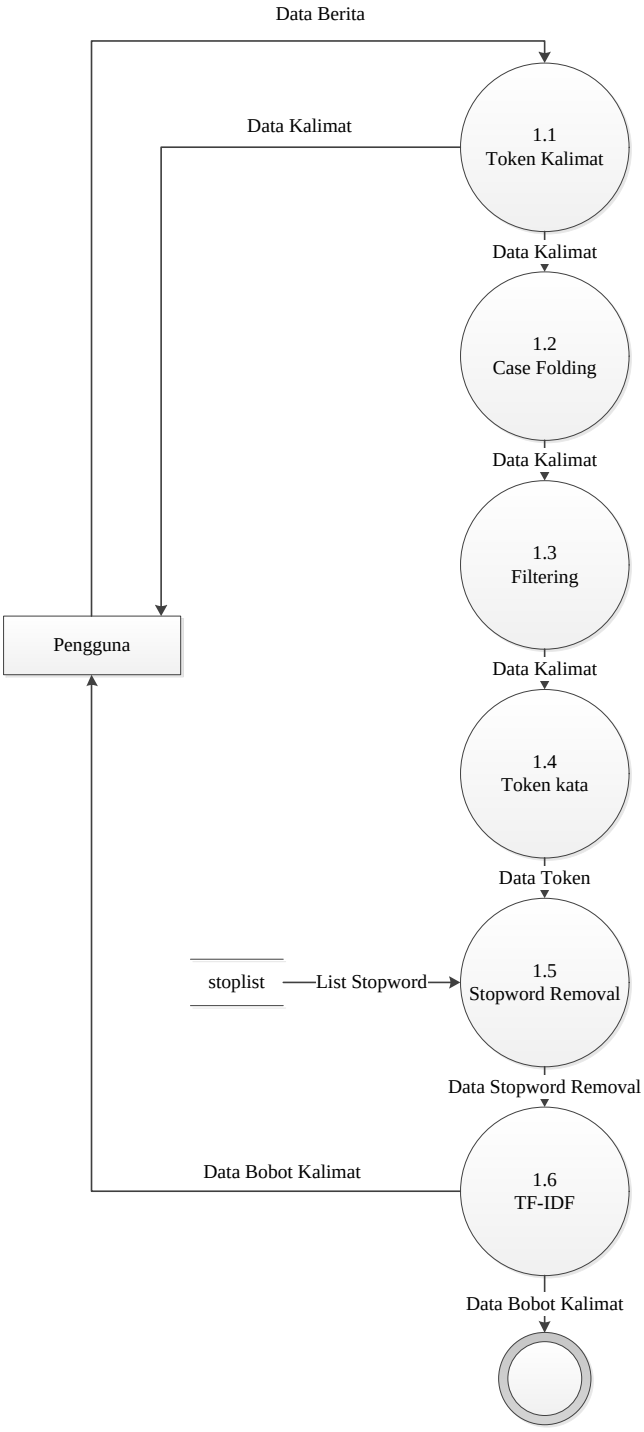
   DFD Level 1 berikut ini menjelaskan bahwa pangguna Aplikasi dapat melakukan akses pada Aplikasi dengan memilih menu yang ada. Untuk lebih jelasnya kita lihat gambar 3.6 berikut.



Gambar 3. 6 DFD Level 1

2. DFD level 2 *Preprocessing*

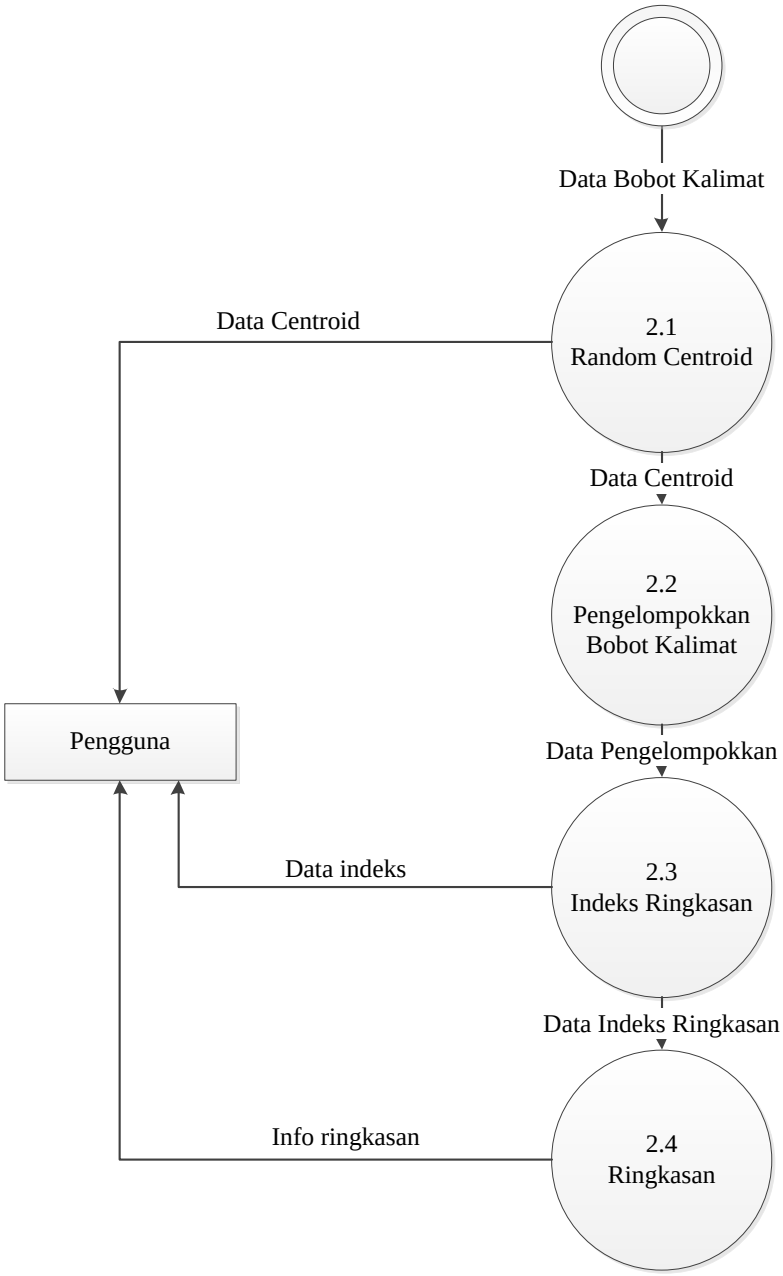
DFD Level 2 berikut ini menjelaskan bahwa data berita untuk kemudian dilakukan *preprocessing*. Untuk lebih jelasnya bisa dilihat pada gambar 3.7 berikut.



Gambar 3. 7 DFD Level 2 *Preprocessing*

3. DFD level 2 *K-Means Clustering*

   DFD level 2 *K-Means Clustering* berikut ini menjelaskan proses pengelompokkan menggunakan metode *K-Means*. Untuk lebih jelasnya bisa dilihat pada gambar 3.8 berikut.



Gambar 3. 8 DFD Level 2 *K-Means Clustering*

3.5.3 Spesifikasi Proses

Spesifikasi proses digunakan untuk menggambarkan proses model aliran yang terdapat pada DFD. Spesifikasi proses dari gambaran DFD pada subbab sebelumnya akan dijelaskan pada tabel berikut ini.

Tabel 3. 21 Spesifikasi Proses

No	Proses	Keterangan
1	No proses	1.1
	Nama proses	Token Kalimat
	Sumber	Pengguna
	Input	Data Berita
	Output	Token Kalimat
	Destination	Pengguna
	Logika user	1. Sistem telah mengambil data berita yang telah dipilih 2. Sistem akan memisahkan suatu teks menjadi sebuah kalimat dengan delimeter titik (“.”)
2	No proses	1.2
	Nama proses	*Case Folding*
	Sumber	Pengguna
	Input	Data Token Kalimat
	Output	Data *Case Folding*
	Destination	Pengguna
	Logika user	1. Sistem mengambil data berita yang telah diubah menjadi per kalimat 2. Sistem akan mengubah semua huruf menadi huruf kecil (*lowcase*)
3	No proses	1.3

	Nama proses	*Filtering*
	Sumber	Pengguna
	Input	Data *Case Folding*
	Output	Data *Filtering*
	Destination	Pengguna
	Logika user	1. Data telah terdiri dari huruf kecil (*lowcase*) 2. Sistem akan menghapus semua simbol dan angka, kecuali huruf, spasi, dan titik (“ ”)
4	No proses	1.4
	Nama proses	Token Kata
	Sumber	Pengguna
	Input	Data *Filtering*
	Output	Data Token Kata
	Destination	Pengguna
	Logika user	1. Sistem telah menghapus semua simbol dan angka, kecuali huruf, spasi, dan titik (“ ”) 2. Sistem akan memisahkan kalimat tersebut menjadi sebuah kata-kata dengan delimeter spasi (“ ”)
5	No proses	1.5
	Nama proses	*Stopword Removal*
	Sumber	Pengguna
	Input	Data Token Kata
	Output	Data *Stopword Removal*
	Destination	Pengguna
	Logika user	1. Sistem telah memisahkan kalimat menjadi kata 2. Sistem akan mengecek kata, apakah kata

		tersebut termasuk ke daftar *stopword* atau tidak.
6	No proses	1.6
	Nama proses	TF-IDF
	Sumber	Pengguna
	Input	Data *Stopword Removal*
	Output	Data Bobot Kalimat
	Destination	Pengguna
	Logika user	1. Sistem telah membagi perkata 2. Sistem akan menghitung bobot dari setiap kata 3. Kemudian sistem akan menjumlahkan bobot setiap kalimat.
7	No proses	2.1
	Nama proses	Random Centroid
	Sumber	Pengguna
	Input	Data Bobot Kalimat
	Output	Data *Centroid*
	Destination	Pengguna
	Logika user	1. Sistem telah menjumlahkan bobot setiap kalimat 2. Sistem akan mengambil nilai secara acak 3. Kemudian sistem akan mengalokasikan ke setiap *cluster*.
8	No proses	2.2
	Nama proses	Pengelompokkan Bobot Kalimat
	Sumber	Pengguna
	Input	Data *Centroid*

	Output	Data Pengelompokkan
	Destination	Pengguna
	Logika user	1. Sistem telah mengalokasikan ke setiap cluster 2. Sistem akan menghitung jarak antara data ke setiap cluster 3. Kemudian sistem akan mengelompokkan data ke setiap *cluster*.
9	No proses	2.3
	Nama proses	Indeks Ringkasan
	Sumber	Pengguna
	Input	Data Pengelompokkan
	Output	Data Indeks Ringkasan
	Destination	Pengguna
	Logika user	1. Sistem telah mengelompokkan ke setiap cluster 2. Sistem mencari indeks kalimat ringkasan
10	No proses	2.4
	Nama proses	Ringkasan
	Sumber	Pengguna
	Input	Data Indeks Ringkasan
	Output	Info ringkasan
	Destination	Pengguna
	Logika user	1. Sistem telah mendapatkan indeks kalimat 2. Sistem menampilkan hasil ringkasan

3.5.4 Kamus Data

Berikut ini adalah kamus data untuk text summarization dengan metode K-Means pada artikel berita berbahasa Indonesia.

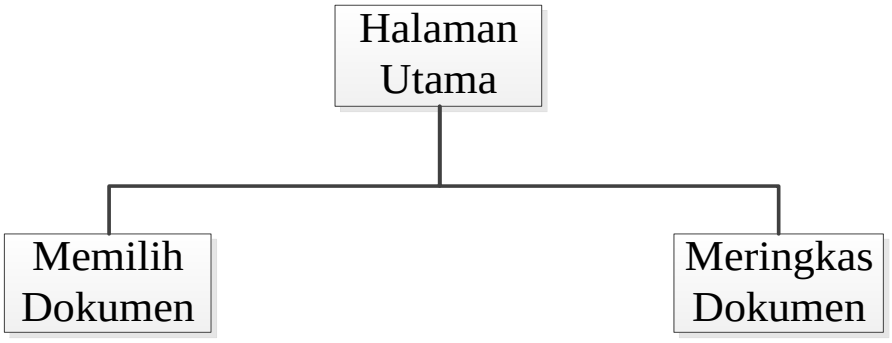
Tabel 3. 22 Kamus Data

1	Nama	Data berita
	Digunakan pada	*Preprocessing* (1.0), token kalimat(1.1)
	Deskripsi	Data berita yang akan diproses dalam peringkasan teks
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
2	Nama	Data Kalimat
	Digunakan pada	*Case folding*(1.2), *filtering*(1.3), token kata(1.4)
	Deskripsi	Data hasil pemrosesan token kalimat,*case folding*, *filtering*
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
3	Nama	Data Token
	Digunakan pada	*Stopword removal* (1.5)
	Deskripsi	Data hasil pemrosesan token kata
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
4	Nama	Data *stopword removal*
	Digunakan pada	TF-IDF(1.6)
	Deskripsi	Data hasil pemrosesan s*topword removal*
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol

5	Nama	Data bobot kalimat
	Digunakan pada	*Random centroid*(2.1)
	Deskripsi	Data hasil pemrosesan tf-idf
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
6	Nama	Data *centroid*
	Digunakan pada	Pengelompokkan bobot kalimat(2.2)
	Deskripsi	Data hasil pemrosesan *random centroid*
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
7	Nama	Data Pengelompokkan
	Digunakan pada	Indeks ringkasan(2.3)
	Deskripsi	Data hasil pemrosesan pengelompokkan bobot kalimat
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
8	Nama	Indeks ringkasan
	Digunakan pada	Ringkasan(2.4)
	Deskripsi	Data hasil data pengelompokkan
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol
9	Nama	Ringkasan
	Digunakan pada	pengguna
	Deskripsi	Data hasil indeks ringkasan
	Struktur data	Teks
	Penjelasan per struktur	[A-Z],[a-z],[0-9], Simbol

3.6 Perancangan Struktur Menu

Srtuktur menu yang digunakan dalam pembangunan text summarization dengan metode K-Means pada artikel berita berbahasa Indonesia dapat dilihat pada gambar berikut.



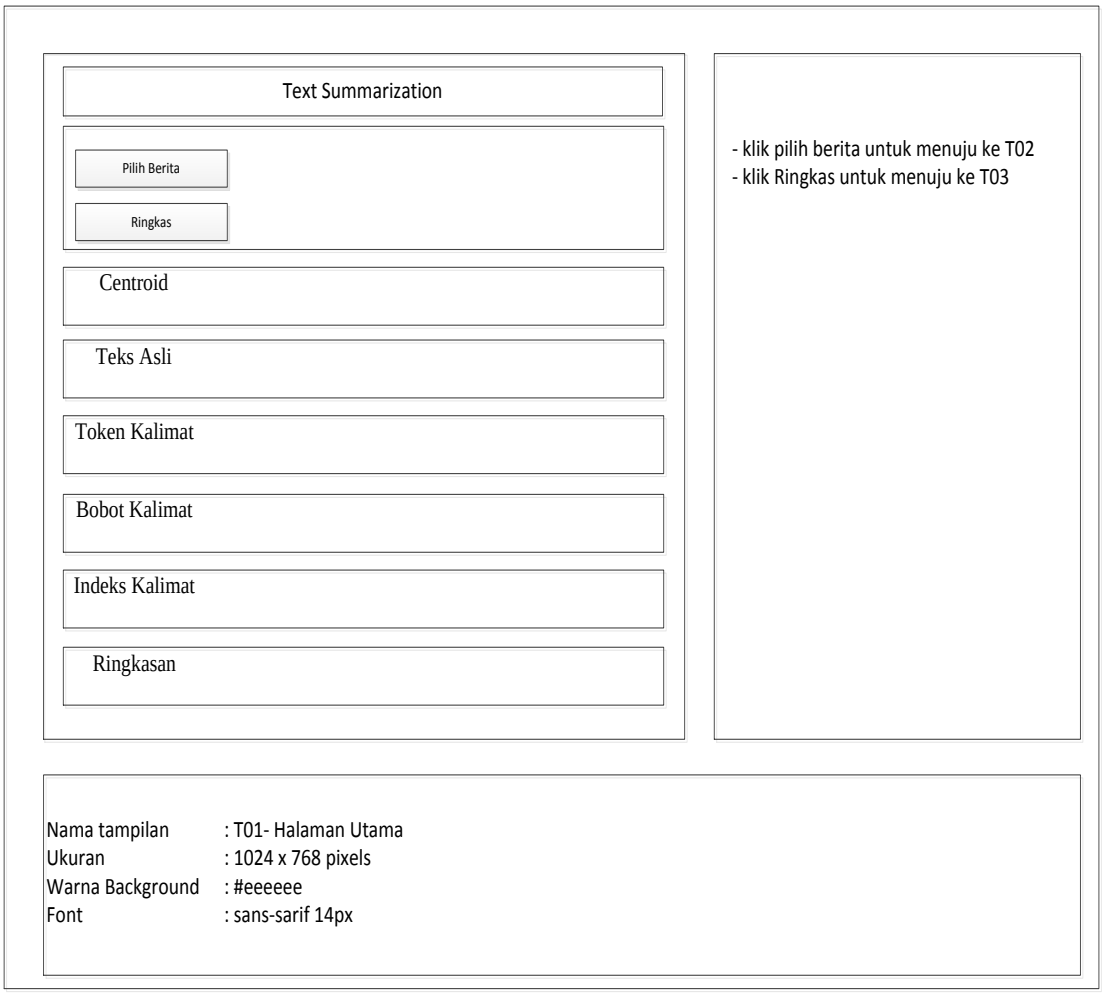
Gambar 3. 9 Struktur menu

3.6.1 Perancangan Antarmuka

Perancangan antarmuka mendeskripsikan rencana tampilan yang akan digunakan pada aplikai yang akan dibangun. Perancangan antarmuka dalam penelitian ini terdiri dari halaman utama, halaman mencari/pilih dokumen, dan halaman hasil ringkasan.

3.6.1.1 Antarmuka Halaman Utama(T01)

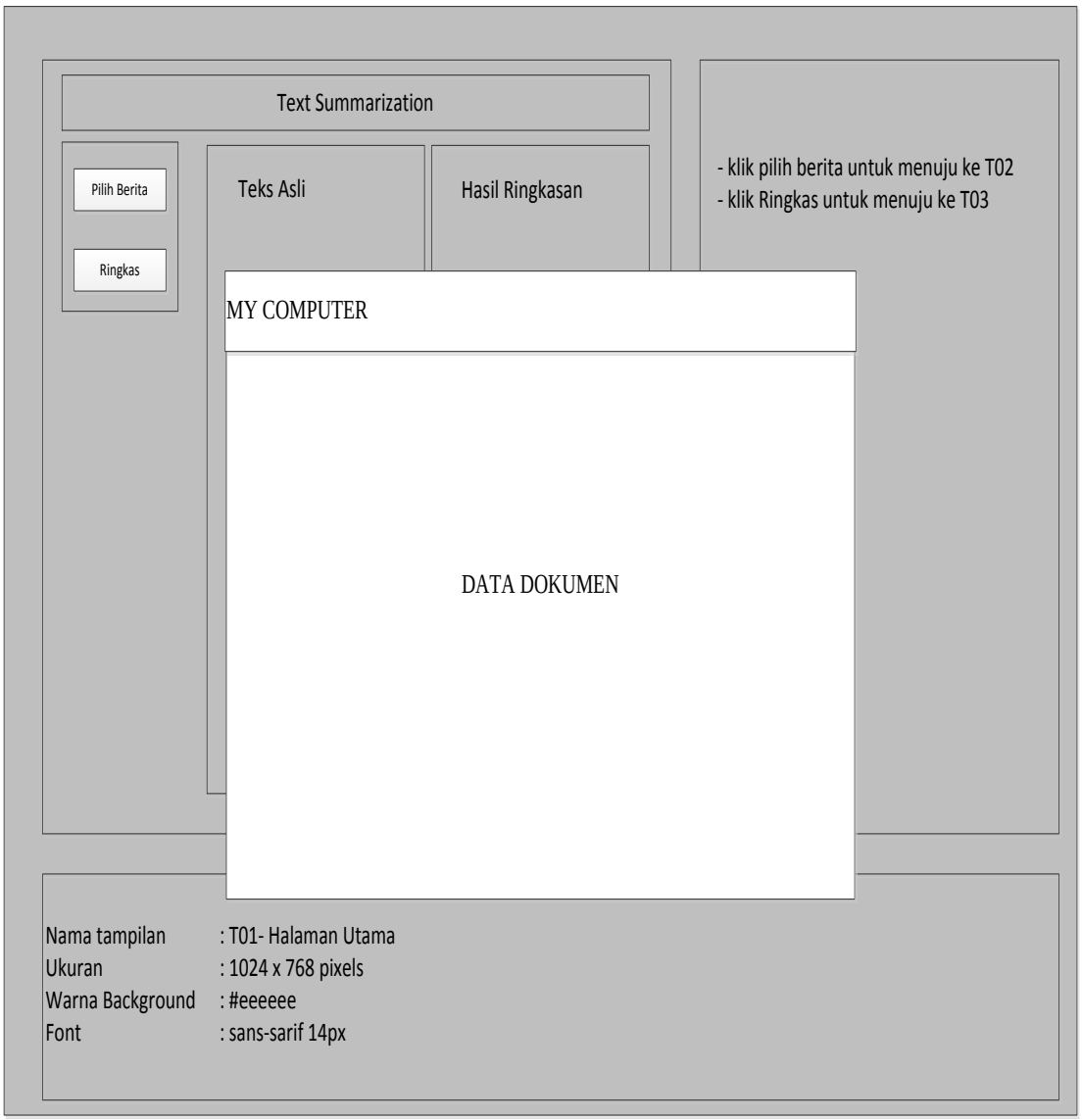
Perancangan halaman Utama(T01) dari sistem peringkasan teks dapat dilihat pada gambar 3.8 berikut.



Gambar 3. 10 Antarmuka Halaman Utama

3.6.1.2 Antarmuka Halaman Mencari/Pilih Dokumen(T02)

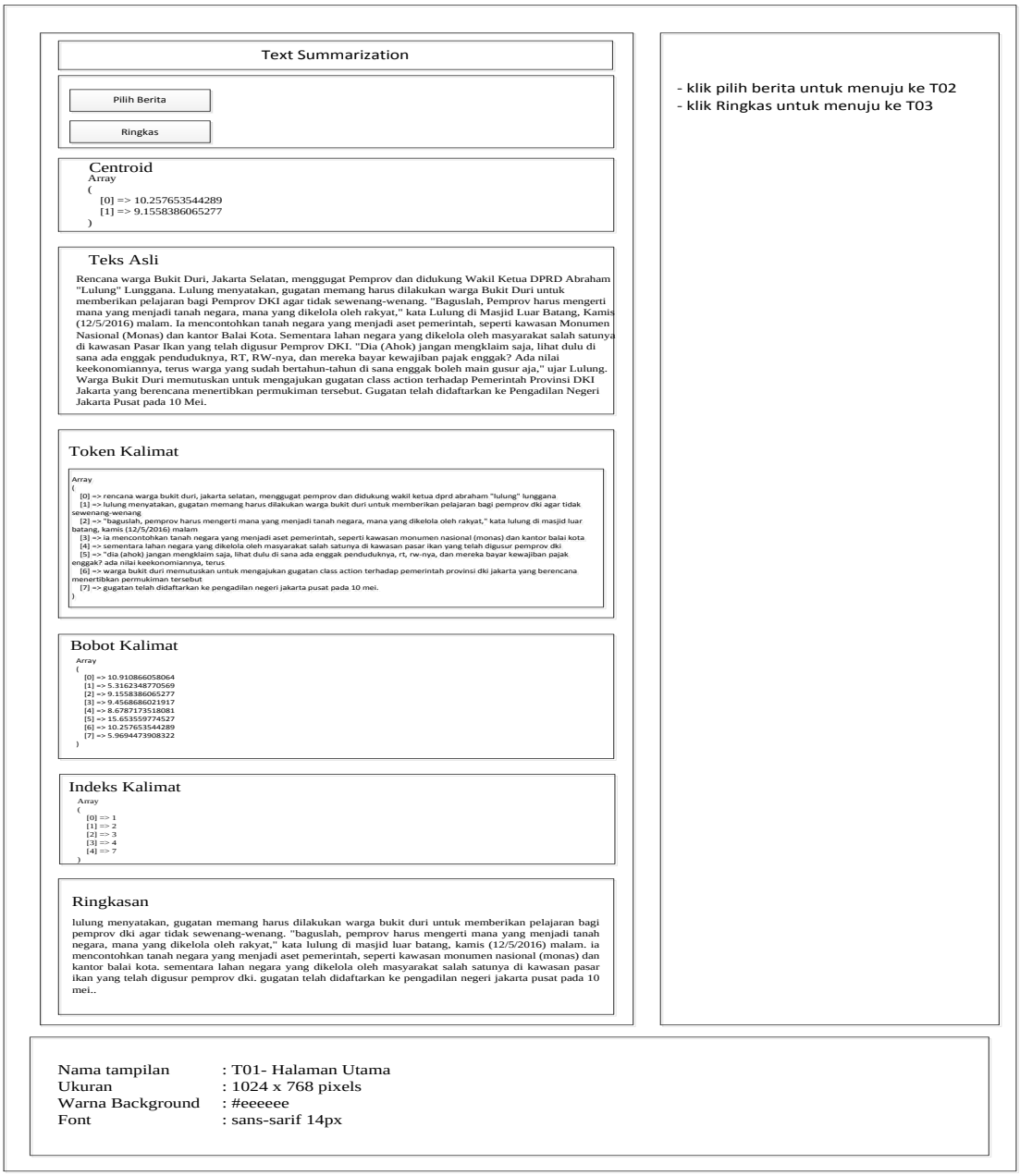
Perancangan halaman mencari/pilih dokumen(T02) dari sistem peringkasan teks dapat dilihat pada gambar 3.9 berikut.



Gambar 3. 11 Antarmuka Halaman Mencari/Pilih Dokumen

3.6.1.3 Antarmuka Halaman Hasil Ringkasan(T03)

Perancangan halaman hasil ringkasan(T03) dari sistem peringkasan teks dapat dilihat pada gambar 3.10 berikut.



Gambar 3. 12 Antarmuka Halaman Hasil Ringkasan

3.6.2 Perancangan Pesan

Berikut perancangan pesan yang terdapat pada aplikasi peringkasan teks otomatis.

3.6.2.1 Perancangan Pesan Berita Kosong(M01)

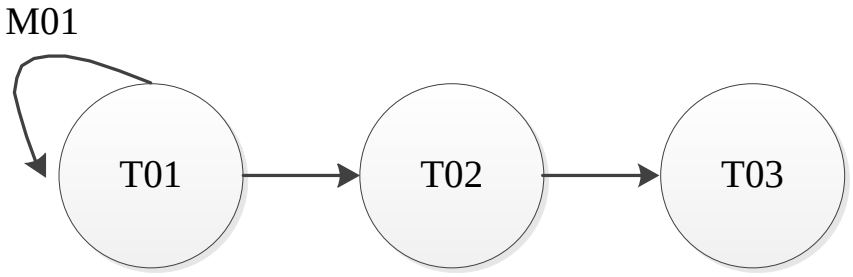
Perancangan pesan berita kosong berisi apabila user belum memilih berita yang akan diringkas. Adapun perancangan pesan berita kosong dapat dilihat pada gamabar 3.11 berikut.



Gambar 3. 13 Perancangan Berita Kosong(M01)

3.6.3 Jaringan Semantik

Jaringan semantik menggambarkan keterhubungan menu dari antarmuka ke antarmuka lain. Jaringan semantik yang terbentuk pada sistem ini dapat dilihat pada gambar 3.11 berikut.



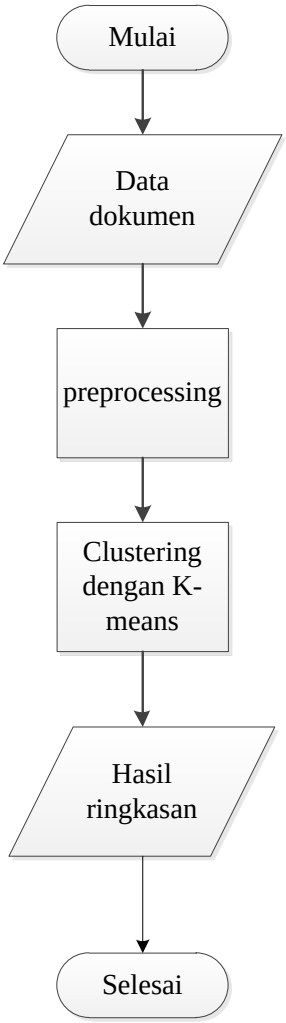
Gambar 3. 14Jaringan Semantik

3.7 Perancangan Prosedural

Perancangan proseduran merupakan flowchart dari preprocessing diantaranya pemecahan kalimat, *case folding, filtering*, *token* kata, *stopword removal*

3.7.1 Perancangan Algoritma Utama

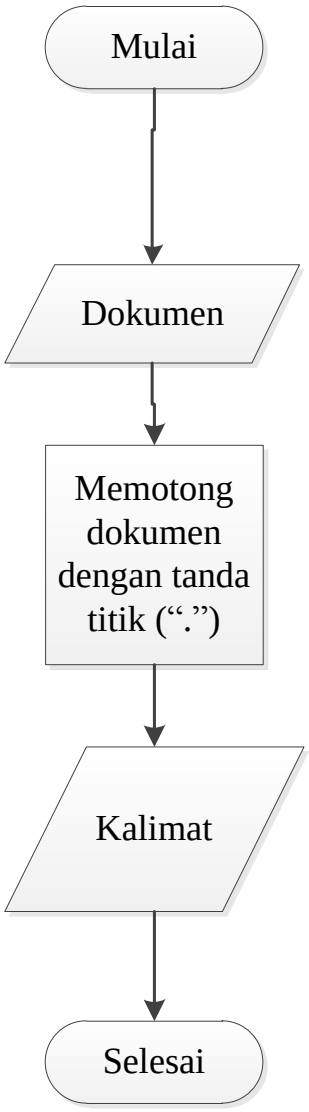
User memilih berita yang akan diringkas. Selanjutnya sistem akan memproses dengan dua tahapan, yaitu preprocessing dan K-Means. Untuk lebih jelasnya dapat diihat pada gambar berikut.



Gambar 3. 15 *Flowchart* Algoritma Utama

3.7.2 Perancangan Algoritma Pemecahan Kalimat

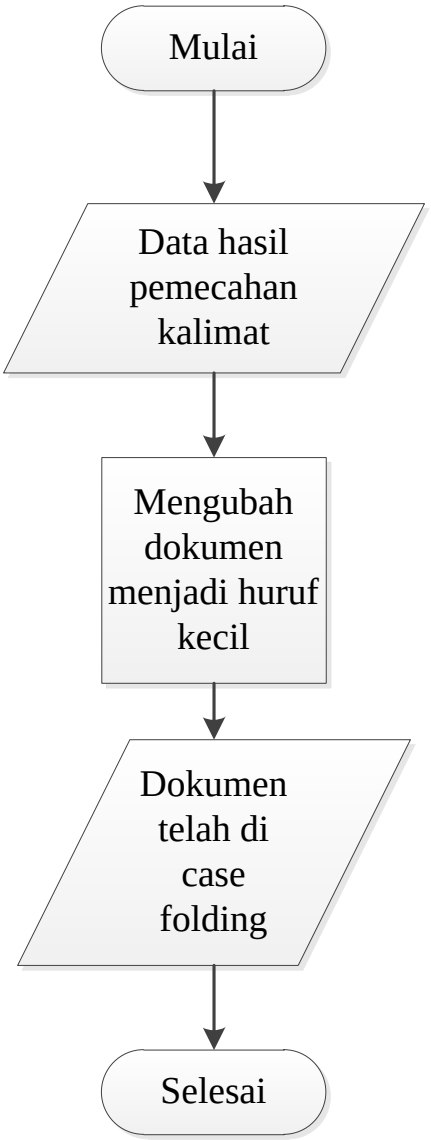
Selanjutnya setelah memilih berita yang akan diringkas, dokumen akan didibagi menjadi perkalimat dengan delimiter titik (“.”). sehingga sistem akan menghasilkan kalimat-kalimat.



Gambar 3. 16 *Flowchart* Algoritma Pemecahan Kalimat

3.7.3 Perancangan Algoritma *Case Folding*

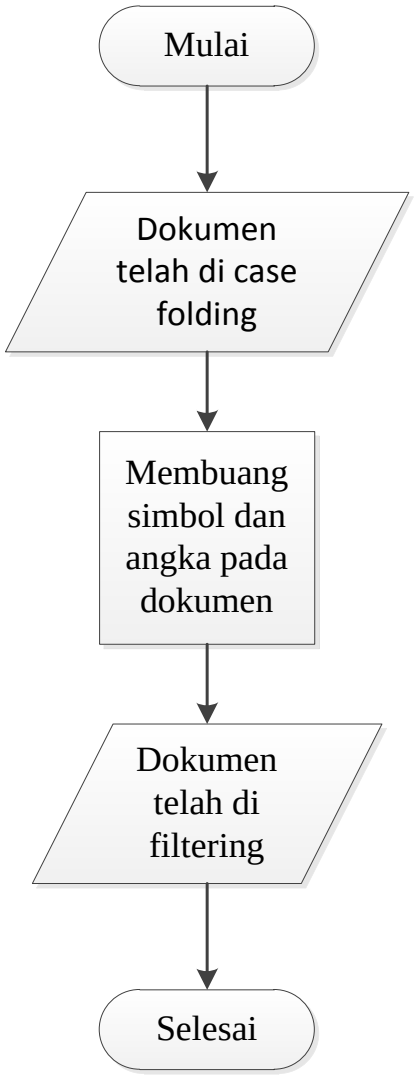
Setelah dibagi menjadi perkalimat, selanjutnya sistem akan merubah semua huruf pada dokumen menjadi huruf kecil.



Gambar 3. 17 *Flowchart* Algoritma *Case Folding*

3.7.4 Perancangan Algoritma *Filtering*

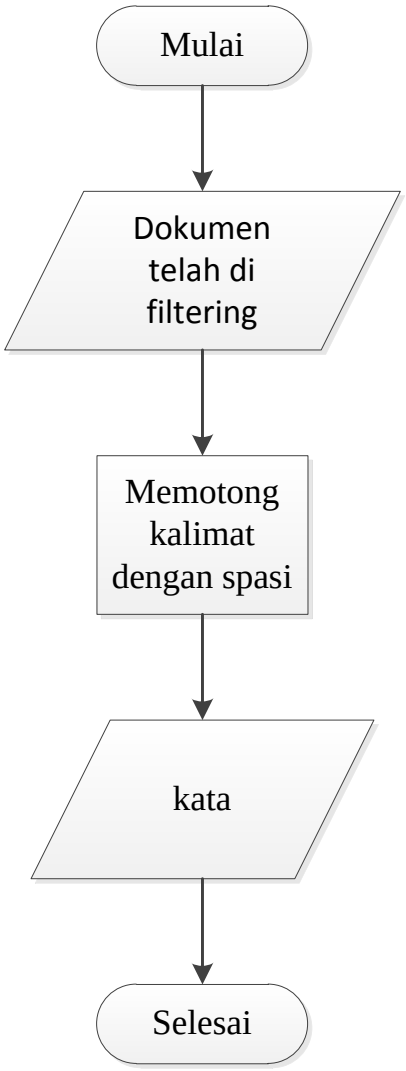
Selanjutnya proses *filtering*, dengan masukkannya dokumen yang telah di *filter* pada proses sebelumnya. Pada proses ini, sistem akan menghapus semua simbol dan angka pada dokumen, kecuali titik. Sehingga sistem akan menghasilkan dokumen yang isinya hanya terdiri dari huruf (A-Z, a-z), spasi (“ ”), dan titik (“.”).



Gambar 3. 18 *Flowchart* Algoritma Filtering

3.7.5 Perancangan Algoritma *Tokenizing*

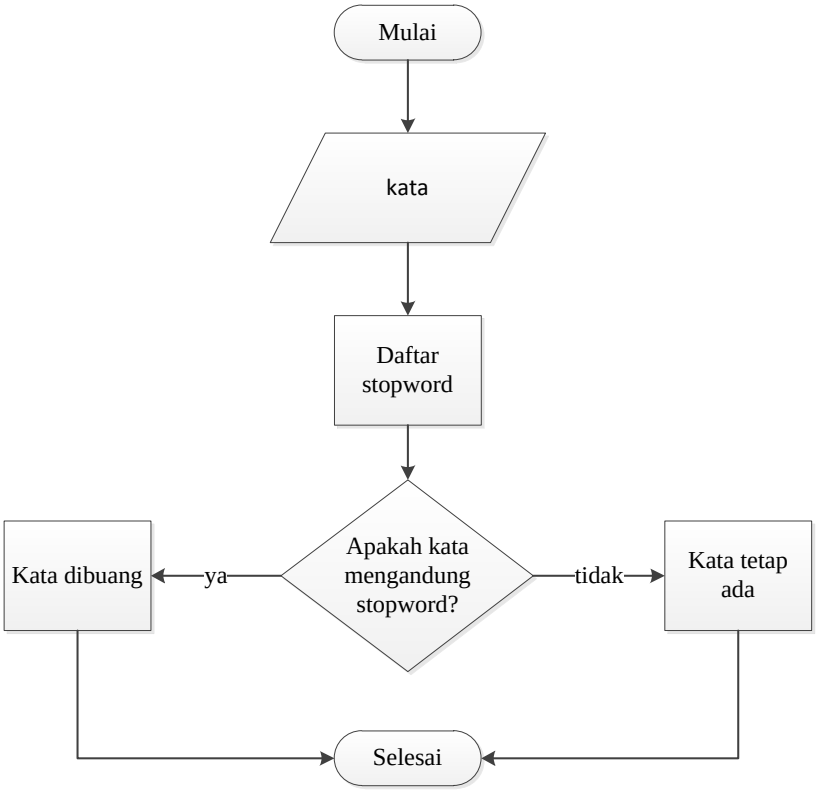
Selanjutnya proses *tokenizing* kata, dengan masukkannya kalimat yang telah diproses sebelumnya. Pada proses ini, sistem akan membagi-bagi kalimat dengan *delimeter* spasi (“ ”) menjadi satuan kata. Sehingga sistem akan menghasilkan kata-kata.



Gambar 3. 19 *Flowchart* Algoritma *Tokenizing*

3.7.6 Perancangan Algoritma *Stopword removal*

Selanjutnya proses *removal stopword*, dengan masukkannya kata yang telah di proses sebelumnya. Pada proses ini, sistem akan mengecek kata dengan daftar *stopword* bisa dilihat pada lampiran daftar *stopword*. Apabila kata tersebut termasuk kedalam daftar *stopword* maka sistem akan membuang kata tersebut. Namun apabila kata tidak termasuk ke dalam daftar *stopword*, maka sistem akan tetap menyimpan kata tersebut, sehingga kata masih tetap ada.



Gambar 3. 20 *Flowchart* Algoritma *Stopword removal*