BAB 3
ANALISIS DAN PERANCANGAN

3.1. Analisis Masalah

Analisis masalah adalah suatu gambaran masalah yang diangkat dalam penulisan skripsi tentang simulasi *Text Summarization* dalam berita bahasa Indonesia. Analisis masalah ini menjelaskan proses identifikasi masalah serta evaluasi mengenai sistem peringkas teks otomatis dalam berita berbahasa Indonesia. Berdasarkan uraian pada latar belakang masalah pada penelitian ini terdapat permasalahan untuk melakukan peringkasan dibutuhkan suatu metode untuk mengatasi masalah redudansi kalimat menggunakan metode *Vektor Space Model* yang nantinya akan terbentuk sebuah ringkasan dari artikel berita berbahasa Indonesia.

3.2. Analisis Sistem

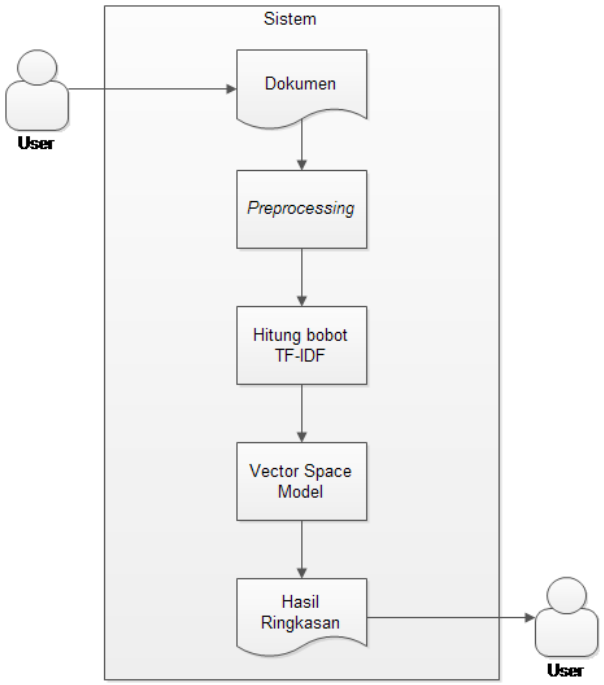
Analisis sistem didefinisikan sebagai penguraian dari sistem utama ke dalam sub-sub sistem dengan tujuan untuk mengidentifikasikan permasalahan permasalahan yang ada dan kebutuhan-kebutuhan yang diperlukan agar dapat diusulkan dan diciptakan sistem yang lebih baik. Dalam mengimplementasikan metode Vector *Space Model* pada Teks *Summarization* akan dibangun dengan gambaran sistem yang akan tertera pada Gambar 3.1.

Analisis sistem peringkas teks otomatis yang dibangun memiliki 3 tahapan yaitu praproses, hitung bobot kata, penentuan ringkasan. Untuk penjelasan setiap tahap pada proses peringkasan dapat dilihat sebagai berikut :

1. Tahap pertama adalah tahap pengumpulan data dokumen berita yang telah dilakukan secara manual dari media online seperti merdeka, republika, liputan6, kompas, tribunnews, dan tempo. Kemudian data tersebut dilakukan proses *preprocessing* yang merupakan tahap dimana data yang akan digunakan sebagai data testing dibersihkan dari *noise* atau dari hal yang tidak mempunyai pengaruh dalam ringkasan. *Preprocessing* yang dilakukan dalam tahap ini

antara lain: pemecahan kalimat, *case folding*, *tokenizing*, hapus *stopword dan stemming.*

2. Tahap kedua adalah melakukan perhitungan bobot menggunakan metode *tf-idf.* dengan menghitung TF(*Term Frequency*), DF(*Document Frequency*), dan dihitung *weight* (W) dari masing-masing term.
3. Tahap ketiga adalah menentukan hasil ringkasan dengan menggunakan metode *Vector Space Model.* Pada metode ini dihitung nilai *similarity* antarkalimat atau kemiripannya pada dokumen masukan. Proses ini dilakukan pada semua kalimat terhadap kalimat yang lainnya yang akan menghasilkan sebuah ringkasan. Hasil ringkasan adalah kumpulan dari kalimat-kalimat hasil persentase yang paling tinggi.



Gambar 3.1 Gambaran Sistem

3.2.1. Analisis Data Masukan (*Input Data*)

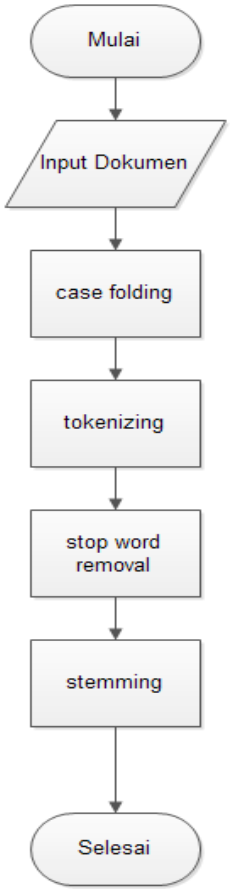
Analisis data pada sistem peringkas teks berita bahasa Indonesia menjelaskan proses data masukan. peringkasan teks otomatis berita yang dibuat

merupakan sistem peringkasan dengan inputan *data training* sekaligus *testing* berupa *single* dokumen untuk menghasilkan ringkasan (*summary*).

Gambaran Sistem menunjukkan skema proses peringkasan secara umum. Berikut alur proses data masukkan peringkasan teks :

1. User memasukkan teks dokumen berupa format teks (file .txt) yang akan diringkas, data masukan tersebut adalah *data training dan testing* dokumen berita dalam berbahasa Indonesia. Proses awal peringkasan *user* memasukkan teks dokumen sesuai Tabel 3.1.
2. Sistem melakukan penyiapan teks (*teks preprocessing*) dokumen yang terdiri dari tahap pemecahan kalimat, *case folding, tokenizing kata, stop word removal* dan *stemming* yang akan dibahas pada analisis berikutnya.

Berikut adalah *flowchart* dari proses *Input Data* dapat dilihat pada gambar 3.2 :



Gambar 3.2 *Flowchart Data Masukan*

Tabel 3.1 Data Training dan Testing

Artikel Berita
Menteri Kelautan dan Perikanan, Susi Pudjiastuti akan melepas semua posisinya di perusahaan penerbangan Susi air. Bahkan agar mencegah conflict of interest, Susi bersedia melepas semua jabatan yang selama ini dipegangnya di sejumlah perusahaan. Yakni, akan melepas jabatan President Direktur PT ASI Pudjiastuti yang bergerak di bidang perikanan dan PT ASI Pudjiastuti Aviation yang jadi operator penerbangan Susi Air. Susi akan melepas semua jabatan CEO Susi air, dan PT ASI grup. “Itu harus dilepaskan, biar bisa kerja maksimal tanpa konflik kepentingan apapun," tuturnya, usai presiden Joko Widodo (Jokowi) mengumumkan kabinet di halaman belakang istana merdeka, jakarta, Minggu (26/19/2014). Yang pasti kerjanya sangat besar. “Tujuh puluh persen wilayah kita lautan," tutur Susi usai diperkenalkan Jokowi di halaman belakang istana merdeka, jakarta, Minggu (26/19/2014). Susi akan pegang teguh pesan Jokowi kepadanya saat dipanggil ke istana merdeka pekan lalu. "Pesannya Pak Jokowi ke saya, Kerja, kerja dan kerja itu saja pegangan saya bekerja," tegasnya. Terkait Program, Susi akan menjawabnya usai pelantikan dan rapat perdana kabinet, Senin (27/10/2014) besok. "Belum tahu programnya, tunggu besok saja selesai pelantikan dan rapat kabinet besok," ujarnya.

Data masukkan yang ada pada Tabel 3.1 didapat dari artikel berita Tribun news dengan judul “Susi Pudjiastuti Langsung Lengser Dari Jabatan Dirut Susi Air” berikut alamat websitenya : http://www.tribunnews.com/nasional/2014/10/26/susi-pudjiastuti-langsung-lengser-dari-jabatan-dirut-susi-air.

3.2.2. Analisis Preprocessing

Analisis preprocessing ini menjelaskan praproses yang dilakukan untuk mendapatkan fitur kata-kata yang bersih sehingga tidak akan mengandung noise yang akan menganggu hasil untuk proses selanjutnya[22]. Berikut adalah tahapan dari praproses sistem peringkas teks berita dalam bahasa Indonesia dapat dilihat pada gambar 3.3.



Gambar 3.3 Tahapan Praproses

A. Pemecahan Kalimat

Pada tahap pemecahan kalimat adalah memecah string dokumen menjadi kumpulan kalimat-kalimat berdasarkan tanda titik " . ",tanda tanya " ? ", dan tanda seru "!" sebagai pemisah *(delimiter)* untuk memotong string dokumen. Tabel 3.2 merupakan hasil proses pemecahan dokumen menjadi kumpulan kalimat-kalimat sebagai berikut :

Tabel 3.2 Pemecahan Kalimat

No	Kalimat
D1	Menteri Kelautan dan Perikanan, Susi Pudjiastuti akan melepas semua posisinya di perusahaan penerbangan Susi air
D2	Bahkan agar mencegah conflict of interest, Susi bersedia melepas semua jabatan yang selama ini dipegangnya di sejumlah perusahaan
D3	Yakni, akan melepas jabatan President Direktur PT ASI Pudjiastuti yang bergerak di bidang perikanan dan PT ASI Pudjiastuti Aviation yang jadi operator penerbangan Susi Air
D4	Susi akan melepas semua jabatan CEO Susi air, dan PT ASI grup

D5	“Itu harus dilepaskan, biar bisa kerja maksimal tanpa konflik kepentingan apapun," tuturnya, usai presiden Joko Widodo (Jokowi) mengumumkan kabinet di halaman belakang istana merdeka, jakarta, Minggu (26/19/2014)
D6	Yang pasti kerjanya sangat besar
D7	“Tujuh puluh persen wilayah kita lautan," tutur Susi usai diperkenalkan Jokowi di halaman belakang istana merdeka, jakarta, Minggu (26/19/2014)
D8	Susi akan pegang teguh pesan Jokowi kepadanya saat dipanggil ke istana merdeka pekan lalu
D9	"Pesannya Pak Jokowi ke saya, Kerja, kerja dan kerja itu saja pegangan saya bekerja," tegasnya
D10	Terkait Program, Susi akan menjawabnya usai pelantikan dan rapat perdana kabinet, Senin (27/10/2014) besok
D11	"Belum tahu programnya, tunggu besok saja selesai pelantikan dan rapat kabinet besok," ujarnya

B. *Case Folding*

Berikut merupakan hasil teks artikel berita yang sudah melalui tahapan pembuangan *case folding:*

Tabel 3.3 *Case Folding*

No	Kalimat
D1	menteri kelautan dan perikanan susi pudjiastuti akan melepas semua posisinya di perusahaan penerbangan susi air
D2	bahkan agar mencegah conflict of interest susi bersedia melepas semua jabatan yang selama ini dipegangnya di sejumlah perusahaan
D3	yakni akan melepas jabatan president direktur pt asi pudjiastuti yang bergerak di bidang perikanan dan pt asi pudjiastuti aviation yang jadi operator penerbangan susi air
D4	susi akan melepas semua jabatan ceo susi air dan pt asi grup
D5	itu harus dilepaskan biar bisa kerja maksimal tanpa konflik kepentingan apapun tuturnya usai presiden joko widodo jokowi mengumumkan kabinet di halaman belakang istana merdeka jakarta minggu
D6	yang pasti kerjanya sangat besar
D7	tujuh puluh persen wilayah kita lautan tutur susi usai diperkenalkan jokowi di halaman belakang istana merdeka jakarta minggu
D8	susi akan pegang teguh pesan jokowi kepadanya saat dipanggil ke istana merdeka pekan lalu
D9	pesannya pak jokowi ke saya kerja kerja dan kerja itu saja pegangan saya bekerja tegasnya
D10	terkait program susi akan menjawabnya usai pelantikan dan rapat perdana kabinet senin besok

D11	belum tahu programnya tunggu besok saja selesai pelantikan dan rapat kabinet besok ujarnya

C. *Tokenizing*

Proses tokenizing kata yaitu menghilangkan karakter pemisah (delimiter) yang menyusunnya berupa karakter spasi. Berikut hasil dari *tokenizing* terdapat pada tabel 3.4.

Tabel 3.4 *Tokenizing*

air	istana	lepaskan	penerbangan	senin
air	jabatan	maksimal	penerbangan	susi
air	jabatan	melepas	perdana	susi
apapun	jabatan	melepas	perikanan	susi
asi	jabatan	melepas	perkenalkan	susi
asi	jakarta	melepas	persen	susi
asi	jakarta	melepas	perusahaan	susi
aviation	joko	mencegah	perusahaan	susi
bekerja	jokowi	mengumumkan	pesan	susi
belum	jokowi	menjawabnya	pesannya	tahu
bersedia	jokowi	menteri	posisinya	tegasnya
besar	jokowi	merdeka	presiden	teguh
besok	kabinet	merdeka	president	terkait
besok	kabinet	merdeka	president	tujuh
besok	kabinet	minggu	program	tunggu
bisa	kelautan	minggu	programnya	tutur
ceo	kepadanya	of	pt	tuturnya
conflict	kepentingan	pak	pt	ujarnya
direktur	kerja	panggil	pudjiastuti	usai
grup	kerja	pegang	pudjiastuti	usai
halaman	kerja	pegangan	puluh	usai
halaman	kerja	pegangnya	rapat	widodo
interest	kerjanya	pekan	rapat	wilayah
istana	konflik	pelantikan	semua	
istana	lautan	pelantikan	semua	

D. *Stop Word Removal*

Proses *Stop Word Removal* dilakukan untuk menghapus kata-kata yang tidak penting. Hasil dari *Tokenizing* kemudian dilakukan proses *stop word*

removal ini. Dan berikut ini adalah hasil dari proses *stop word removal* pada tabel 3.5.

Tabel 3.5 *Stop Word Removal*

air	jabatan	mencegah	perikanan	senin
apapun	jakarta	mengumumkan	perkenalkan	susi
asi	joko	menjawabnya	persen	tegasnya
aviation	jokowi	menteri	perusahaan	teguh
bekerja	kabinet	merdeka	pesan	terkait
bersedia	kelautan	minggu	pesannya	tujuh
besar	kepadanya	of	posisinya	tunggu
besok	kepentingan	panggil	presiden	tutur
ceo	kerja	pegang	president	tuturnya
conflict	kerjanya	pegangan	program	ujarnya
direktur	konflik	pegangnya	programnya	usai
grup	lautan	pekan	pt	wilayah
halaman	lepaskan	pelantikan	pudjiastuti	widodo
interest	maksimal	penerbangan	puluh	
istana	melepas	perdana	rapat	

E. *Stemming*

Proses stemming yang dilakukan adalah untuk menentukan kata dasar yang belum diproses melalui stopword dan tidak terdapat pada database. Hasil dari *stop word removal* kemudian dilakukan proses *stemming* ini. Dan berikut ini adalah hasil dari proses *stemming* pada tabel 3.6.

Tabel 3.6 *Stemming*

air	jabat	cegah	ikan	senin
apa	jakarta	umum	kenal	susi
asi	joko	jawab	persen	tegas
aviation	jokowi	menteri	usaha	teguh
kerja	kabinet	merdeka	pesan	terkait
sedia	laut	minggu	pesan	tujuh
besar	kepada	of	posisi	tunggu
besok	penting	panggil	presiden	tutur
ceo	kerja	pegang	president	tutur
conflict	kerja	pegang	program	ujar
direktur	konflik	pegang	program	usai
grup	lautan	pekan	pt	wilayah

halaman	lepas	lantik	pudjiastuti	widodo
interest	maksimal	terbang	puluh	jumlah
istana	lepas	dana	rapat	

Berikut adalah kumpulan kalimat hasil stemming yang akan dihitung bobot tiap kata terhadap setiap kalimatnya yang akan diproses pada algoritma *TF-IDF* dapat dilihat pada Tabel 3.7.

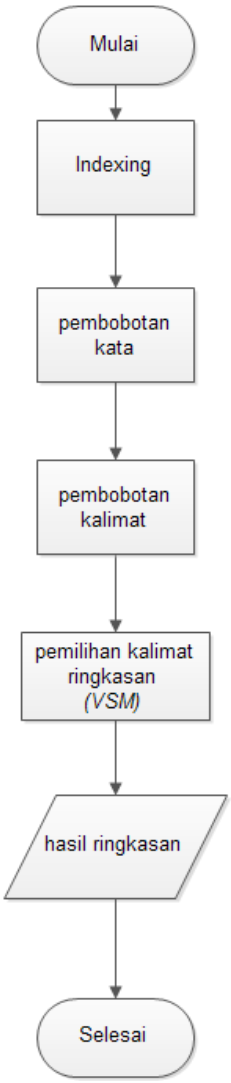
Tabel 3.7 Kalimat Hasil Stemming

No	Kalimat
D1	menteri laut ikan susi pudjiastuti lepas posisi usaha terbang susi air
D2	cegah conflict of interest susi sedia lepas jabat pegang jumlah usaha
D3	lepas jabatan president direktur pt asi pudjiastuti gerak bidang ikan pt asi pudjiastuti aviation operator terbang susi air
D4	susi lepas jabat ceo susi air dan pt asi grup
D5	lepas kerja maksimal konflik penting apa tutur usai presiden joko widodo jokowi umumk kabinet halaman belakang istana merdeka jakarta minggu
D6	kerja besar
D7	tujuh puluh persen wilayah laut tutur susi usai kenal jokowi halaman belakang istana merdeka jakarta minggu
D8	susi pegang teguh pesan jokowi kepada panggil istana merdeka pekan lalu
D9	pesan jokowi kerja kerja kerja pegang kerja tegas
D10	terkait program susi jawab usai lantik rapat dana kabinet senin besok
D11	program tunggu selesai lantik rapat kabinet ujarn

3.2.3. Analisis Metode Summarization

Pada analisis metode *summarization* akan menjelaskan proses peringkasan teks pembobotan kata dengan TF-IDF dan *Vektor Space Model*. Sebelumnya akan dilakukan terlebih dahulu proses pemilihan kata memilih kata berdasarkan bobot kata tersebut dalam dokumen.

Berikut adalah *flowchart* proses retrieval menggunakan *Vector Space Model* dapat dilihat pada gambar 3.4.



Gambar 3.4 *Flowchart Retrieval* Menggunakan *Vector Space Model*

A. Algoritma *Tf-Idf*

Tahapan ini meliputi identifikasi faktor yang sangat penting tentang apa yang dibicarakan dalam teks tersebut. Adapun teknik yang digunakan yaitu teknik *word frequency*. Teknik ini mendeteksi topik berdasarkan kata-kata yang sering muncul dalam suatu teks dokumen [22].

Pemberian bobot pada masing-masing *term* didefinisikan melalui perhitungan frekuensi kemunculan dokumen yang mengandung sebuah *term* yaitu DF (*document frequency*), penghitungan frekuensi kemunculan *term* di dalam dokumen yaitu TF (*term frequency)* dan perhitungan jumlah dokumen yang

mengandung sebuah *term* yang dicari dari kumpulan dokumen yang ada yaitu IDF *(Inverse Document Frequency)* yang mengacu ke rumus (2). Hasil perhitungan nilai TF-IDF pada term tiap dokumen pada data *training* dan *testing* dapat dilihat pada tabel 3.8 dan table 3.9:

Keterangan

Kata/Term : Daftar kata/*term* yang terdapat pada data training dan testing

D: Dokumen/Kalimat

df: Frekuensi kata pada dokumen

IDF: bobot keanggotaan kata pada dokumen

Tabel 3.8 Perhitungan Nilai TF-IDF Pada *Data Training dan Testing*

Kata yang dianggap penting (query)	tf											df	D/df	IDF log (D/df)
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11			
menteri	1	0	0	0	0	0	0	0	0	0	0	1	11	1.041
laut	1	0	0	0	0	0	0	0	0	0	0	1	11	1.041
ikan	1	0	1	0	0	0	0	0	0	0	0	2	5.5	0.740
lepas	1	1	1	1	0	0	0	0	0	0	0	4	2.75	0.439
jabat	1	1	1	1	0	0	0	0	0	0	0	4	2.75	0.439
usaha	1	1	0	0	0	0	0	0	0	0	0	2	5.5	0.740
terbang	1	1	0	0	0	0	0	0	0	0	0	2	5.5	0.740
umum	0	0	0	0	1	0	0	0	0	0	0	1	11	1.041
kerja	0	0	0	0	0	1	0	0	0	0	0	1	11	1.041

Tabel 3.9 Perhitungan Bobot Nilai (W) Pada *Data Training dan Testing*

Kata yang dianggap penting (query)	W										
	D1 * IDF	D2 * IDF	D3 * IDF	D4 * IDF	D5 * IDF	D6 * IDF	D7 * IDF	D8 * IDF	D9 * IDF	D10 * IDF	D11 * IDF
menteri	1.041	0	0	0	0	0	0	0	0	0	0
laut	1.041	0	0	0	0	0	0	0	0	0	0
ikan	0.740	0	0.740	0	0	0	0	0	0	0	0
lepas	0.439	0.439	0.439	0.439	0	0	0	0	0	0	0
jabat	0.439	0.439	0.439	0.439	0	0	0	0	0	0	0
usaha	0.740	0.740	0	0	0	0	0	0	0	0	0
terbang	0.740	0.740	0	0	0	0	0	0	0	0	0
umum	0	0	0	0	1.041	0	0	0	0	0	0
kerja	0	0	0	0	0	1.041	0	0	0	0	0

.

Dari studi kasus diatas dapat diketahui bahwa nilai bobot (W) dari D1, D2, D3, dan D4 adalah sama. Apabila hasil bobot pada kalimat tidak dapat diurutkan secara tepat karena nilai W keduanya sama, maka diperlukan proses perhitungan dengan algoritma *vector space model.* Intisari dari metode ini adalah menghitung nilai cosinus sudut dari dua vector, yaitu nilai W dari tiap kalimat dan W dari kata kunci(*query)* untuk mendapatkan kalimat yang lebih efektif untuk dijadikan ringkasan.

B. Algoritma *Vector Space Model*

Vector Space Model adalah metode untuk menggambarkan fitur pada text dalam bentuk vektor. Setiap kalimat dianggap sebuah vektor dengan nilai berupa keanggotaan atau bobot kalimat yang didapat dari proses perhitungan TF-IDF. Untuk memilih kalimat paling sesuai maka digunakan perhitungan cosines similarity dengan query daftar kata penting.

Formula yang digunakan pada VSM dan diimplemetasikan pada tabel 3.10 dan tabel 3.11:

$Sqrt(kk) = Sqrt \sum_{j=1}^{n} kk_j^2$ **(3.1)**
Dimana :
j= kata di basis data
n= jumlah kata di basis data
kk= kata kunci

$Sqrt(D_i) = Sqrt(\sum_{j=1}^{n} D_{i,j}^2)$ **(3.2)**
Dimana :
j= kata di basis data
n= jumlah kata di basis data
i= dokumen ke-i
D = urutan dokumen(kalimat) ke-I untuk kata ke-j pada basis data

Untuk
$Sum(kk\ dot\ D_i) = kk_j D_{i,j}$**(3.3)**
Dimana :
j= kata di basis data
n= jumlah kata di basis data
i= dokumen ke-i
kk= kata kunci ke-j dari basis data
D= urutan dokumen(kalimat) ke-i untuk kata ke-j pada basis data

Tabel 3.10 Perhitungan *Vector Space Model* dari Hasil *TF-IDF* Pada *Data Training dan Testing* (1)

Kata	Query (KK)	W1^2 (D1)	W2^2 (D2)	W3^2 (D3)	W4^2 (D4)	W5^2 (D5)	W6^2 (D6)	W7^2 (D7)	W8^2 (D8)	W9^2 (D9)	W10^2 (D10)	W11^2 (D11)
menteri	0	1.083	0	0	0	0	0	0	0	0	0	0
laut	0	1.083	0	0	0	0	0	0	0	0	0	0
ikan	0	0.547	0	0.547	0	0	0	0	0	0	0	0
lepas	0.192	0.192	0.192	0.192	0.192	0	0	0	0	0	0	0
jabat	0.192	0.192	0.192	0.192	0.192	0	0	0	0	0	0	0
usaha	0	0.547	0.547	0	0	0	0	0	0	0	0	0
terbang	0	0.547	0.547	0	0	0	0	0	0	0	0	0
umum	0	0	0	0	0	1.083	0	0	0	0	0	0
kerja	0	0	0	0	0	0	1.083	0	0	0	0	0
Jumlah	**0.384**	**4.191**	**1.478**	**0.931**	**0.384**	**1.083**	**1.083**	**0**	**0**	**0**	**0**	**0**
Sqrt	**0.619**	**2.047**	**1.215**	**0.964**	**0.619**	**1.040**	**1.040**	**0**	**0**	**0**	**0**	**0**

Tabel 3.11 Perhitungan *Vector Space Model* dari Hasil *TF-IDF* Pada *Data Training dan Testing* (2)

Kata	Query (KK)	KK*D1	KK*D2	KK*D3	KK*D4	KK*D5	KK*D6	KK*D7	KK*D8	KK*D9	KK*D10	KK*D11
menteri	0	0	0	0	0	0	0	0	0	0	0	0
laut	0	0	0	0	0	0	0	0	0	0	0	0
ikan	0	0	0	0	0	0	0	0	0	0	0	0
lepas	0.192	0.036	0.036	0.036	0.036	0.036	0	0	0	0	0	0
jabat	0.192	0.036	0.036	0.036	0.036	0.036	0	0	0	0	0	0
usaha	0	0	0	0	0	0	0	0	0	0	0	0
terbang	0	0	0	0	0	0	0	0	0	0	0	0
umum	0	0	0	0	0	0	0	0	0	0	0	0
kerja	0	0	0	0	0	0	0	0	0	0	0	0
Jumlah	**0.384**	**0.072**	**0.072**	**0.072**	**0.072**	**0.072**	**0**	**0**	**0**	**0**	**0**	**0**
Sqrt	**0.619**	**0.268**	**0.268**	**0.268**	**0.268**	**0.268**	**0**	**0**	**0**	**0**	**0**	**0**

Rasio kompresi *(Compression Rate)* pada suatu ringkasan berfungsi untuk menentukan persentasi batas panjang ringkasan yang akan ditampilkan [22]. *Compression rate* pada proses peringkasan akan menentukan panjang ringkasan yang dihasilkan. Hasil ringkasan dengan memilih *compression* / batas panjang ringkasan 35%.

Untuk

$$cosine\ \mathrm{Di} = \sum_{j=1}^{n}(\mathrm{kkj} * \mathrm{Di,j}) \Big/ (\mathrm{sqrt(kk)} * \mathrm{sqrt(Di)}) \quad \textbf{(3.4)}$$

Dimana:

j= kata di basis data

n= jumlah kata di basis data

i= dokumen ke-i

kk= kata kunci ke-j dari basis data

D= urutan dokumen(kalimat) ke-i untuk kata ke-j pada basis data

1. Untuk kalimat 1 (D1)
   Cosine (D1) = sum (kk . D1) / (sqrt(kk) * sqrt(D1)
   = 0.268 / (0.619 * 2.047)
   = 0.211

2. Untuk kalimat 2 (D2)
   Cosine (D2) = sum (kk . D2) / (sqrt(kk) * sqrt(D2)
   = 0.268/ (0.619 * 1.215)
   = 0.356 **(ringkasan)**

3. Untuk kalimat 3 (D3)
   Cosine (D3) = sum (kk . D3) / (sqrt(kk) * sqrt(D3)
   = 0.268/ (0.619 * 0.964)
   = 0.449 **(ringkasan)**

4. Untuk kalimat 4 (D4)
   Cosine (D4) = sum (kk . D4) / (sqrt(kk) * sqrt(D4)
   = 0.268/ (0.619 * 0.619)
   = 0.699 **(ringkasan)**

5. Untuk kalimat 5 (D5)
   Cosine (D5) = sum (kk . D5) / (sqrt(kk) * sqrt(D5)
   = 0.268/ (0.619 * 1.040)
   = 0.416 **(ringkasan)**

6. Untuk kalimat 6 (D6)
   Cosine (D6) = sum (kk . D6) / (sqrt(kk) * sqrt(D6)
   = 0 / (0.619 * 1.040)
   = 0

7. Untuk kalimat 7 (D7)
   Cosine (D7) = sum (kk . D7) / (sqrt(kk) * sqrt(D7)
   = 0 / (0.619 * 0)
   = 0
8. Untuk kalimat 8 (D8)
   Cosine (D8) = sum (kk . D8) / (sqrt(kk) * sqrt(D8)
   = 0 / (0.619 * 0)
   = 0

9. Untuk kalimat 9 (D9)
   Cosine (D9) = sum (kk . D9) / (sqrt(kk) * sqrt(D9)
   = 0 / (0.619 * 0)
   = 0

10. Untuk kalimat10 (D10)
    Cosine (D10) = sum (kk . D10) / (sqrt(kk) * sqrt(D10)
    = 0 / (0.619 * 0)
    = 0

11. Untuk kalimat 11 (D11)
    Cosine (D11) = sum (kk . D11) / (sqrt(kk) * sqrt(D11)
    = 0 / (0.619 * 0)
    = 0

Berikut adalah urutan dokumen hasil ringkasan dengan nilai kompresi 35%. Maka 35% dari 11 Dokumen yaitu : 11 x 35% = 3.85 ≈ 4 Dokumen.

Tabel 3.12 Urutan Dokumen Hasil Ringkasan 35%

Dokumen	Bobot
D2	0.356
D3	0.449
D4	0.699
D5	0.416

Sesuai perhitungan diatas maka nilai cosinus setiap dokumen telah didapat dari hasil akhir perhitungan *vector space model* diketahui bahwa kalimat yang diambil untuk dijadikan sebuah ringkasan terdapat pada kalimat ke 2, 3, 4, dan 5 dengan *compression rate* 35%. Semakin besar hasil cosinus tersebut maka semakin

Tabel 3.13 Hasil Ringkasan Teks Artikel *Compression* 35%

No	Kalimat
D2	Bahkan agar mencegah conflict of interest, Susi bersedia melepas semua jabatan yang selama ini dipegangnya di sejumlah perusahaan
D3	Yakni, akan melepas jabatan President Direktur PT ASI Pudjiastuti yang bergerak di bidang perikanan dan PT ASI Pudjiastuti Aviation yang jadi operator penerbangan Susi Air
D4	Susi akan melepas semua jabatan CEO Susi air, dan PT ASI grup
D5	“Itu harus dilepaskan, biar bisa kerja maksimal tanpa konflik kepentingan apapun," tuturnya, usai presiden Joko Widodo (Jokowi) mengumumkan kabinet di halaman belakang istana merdeka, jakarta, Minggu (26/19/2014)

3.3. Spesifikasi Kebutuhan Perangkat Lunak

Spesifikasi kebutuhan perangkat lunak yang dibangun dibagi menjadi dua kebutuhan, yaitu kebutuhan nonfungsional dan fungsional. Spesifikasi kebutuhan fungsional dan nonfungsional dapat dilihat pada tabel 3.14.

Tabel 3.14 SKPL

Kode	Keterangan
Fungsional	
SKPL – F – 001	Aplikasi ini dapat membantu pengguna dalam membaca intisari dari dokumen / ringkasan
SKPL – F – 002	Aplikasi ini dapat melakukan preprocessing untuk mengubah data mentah menjadi siap pakai
SKPL – F – 003	Aplikasi ini dapat membatu pengguna mempersingkat waktu membaca dalam menentukan intisari dokumen
Non Fungsional	
SKPL – NF – 001	Sistem yang dibangun berbasis web
SKPL – NF – 002	Aplikasi yang dibangun dapat berjalan optimal dibrowser google chrome dan mozila firefox
SKPL – NF – 003	Aplikasi ini dibangun dengan spesifikasi hardware yang memenuhi standar minimum kebutuhan.
SKPL – NF – 004	Data yang dipakai untuk ringkasan dokumen adalah data mentah yang dimbil dari media online

	seperti merdeka, republika, liputan6, kompas, tribunnews, dan tempo.
SKPL – NF – 005	Data dokumen yang diinputan adalah format file text (.txt)

3.4. Analisis Kebutuhan Nonfungsional

Analisis kebutuhan *Non Fungsional* menggambarkan kebutuhan luar sistem yang diperlukan untuk menjalankan perangkat lunak yang dibangun. Analisis kebutuhan non fungsional yang dilakukan diantaranya Analisis Kebutuhan Perangkat Lunak (*Software*), dan Analisis Kebutuhan Perangkat Keras (*Hardware*).

3.4.1. Analisis Kebutuhan Perangkat Keras (*Hardware*)

Kebutuhan Perangkat keras yang mendukung dalam pembangunan sistem peringkas teks otomatis ini adalah sebagai berikut :

1. *Processor* amd 1,40 GHz
2. *RAM* 2048MB
3. Monitor dengan resolusi 1366 x 768 pixels
4. Keyboard
5. Mouse

3.4.2. Analisis Kebutuhan Perangkat Lunak (*Software*)

Analisis kebutuhan perangkat lunak (*software*) pada penelitian ini yang mendukung dalam sistem peringkas teks otomatis ini adalah sebagai berikut :

1. Sistem Operasi Window 8.1 Profesional 64 bit.
2. Bahasa Pemrograman PHP
3. Web server XAMPP v3.2.1
4. *Code Editor* SublimeText 3 Portable
5. Adobe Photoshop CS6
6. Web Browser Google Chrome 43.0.2357.124 m

3.4.3. Analisis Kebutuhan Perangkat Pikir (*Brainware*)

Kebutuhan perangkat pikir (*Brainware* yang dibahas dalam penelitian ini dikelompokkan menjadi tiga kategori yaitu *user knowledge and experience*, *user jobs and tasks*, dan *user physical characteristic*. Penjelasan untuk masing-masing kategori sebagai berikut:

1. Pengetahuan dan Pengalaman Pengguna (*User Knowledge and Experience)*

   Pengetahuan dan pengalaman pengguna yang di targetkan dalam pembangunan perangkat lunak ini dapat dilihat pada table 3.16

Tabel 3.16 User Knowledge and Experience

Aspek	Pengguna
Education Level	SMA sampai Perguruan tinggi (D3,S1,S2,S3)
Reading Level	Sedang Sampai Tinggi
Computer Literacy	*moderate experience till highly technical*
Task Experience	Rendah sampai Tinggi
System Experience	Rendah sampai Tinggi
Application Experience	Rendah sampai Tinggi
Use Of Other System	No
Native Language or Culture	Dominan Indonesia

2. Tugas dan Kebutuhan Pengguna (*User Jobs and Tasks*)

   Tugas dan pekerjaan pengguna yang ditargetkan dalam pembangunan perangkat lunak ini dapat dilihat pada table 3.17

Tabel 3.17 Tugas dan Kebutuhan Pengguna

Aspek	Pengguna
Frequency Of Use	*Continual*
Primary Training	Disediakan menu petunjuk
Task or Need Importance	Sedang sampai Tinggi
Task Structure	Sedang sampai Tinggi
Job Categories	No
Life Style	No

3. Karakteristik Pengguna (*User Physical Characteristic*)

   Karakter fisik pengguna yang ditargetkan dalam pembangunan perangkat lunak ini dapat dilihat pada table 3.18

Tabel 3.18 Karakteristik Pengguna

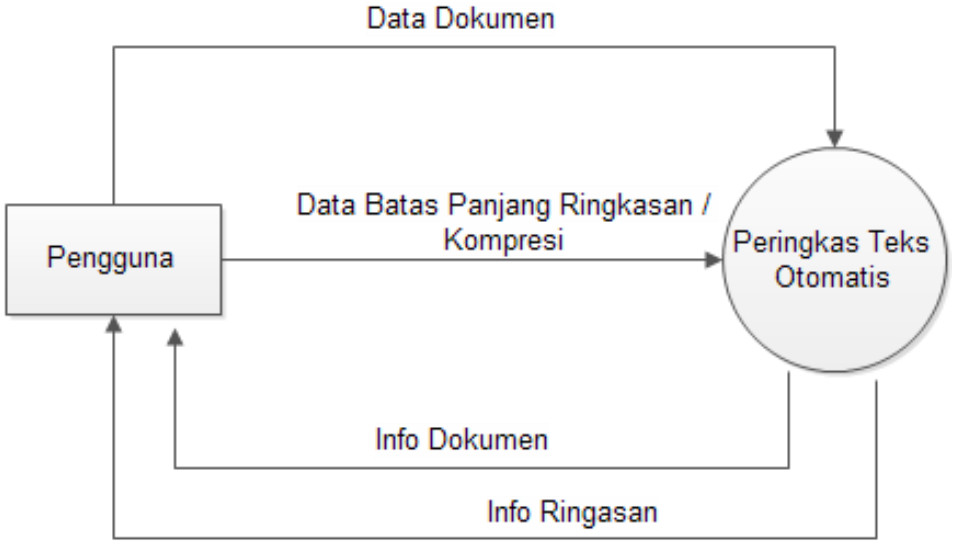
Aspek	Pengguna
Handedness	Keduanya dalam kondisi baik
Gender	Laki-laki dan perempuan
Age	20 tahun ke atas
Disabilities	Tidak cacat tubuh

3.5. Analisis 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, dan spesifikasi proses.

3.5.1. Diagram Konteks

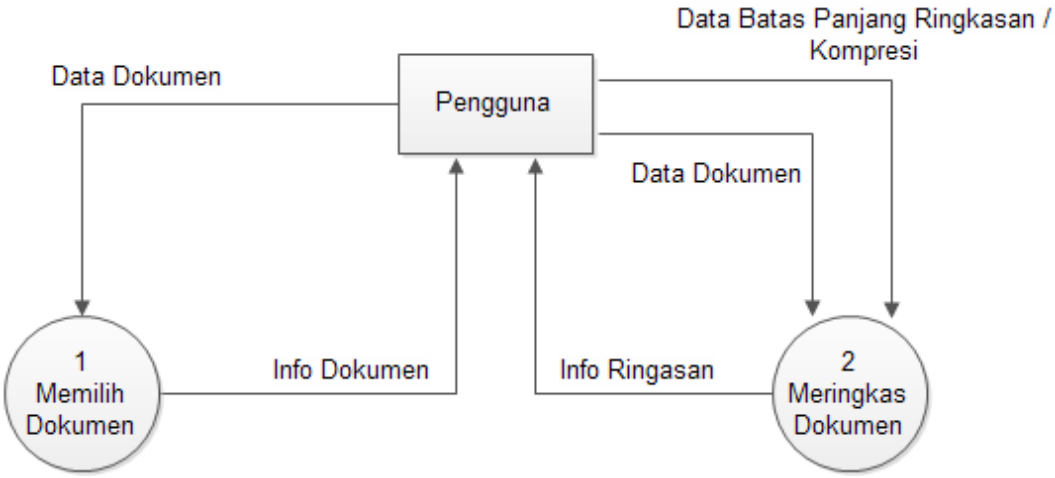
Diagram konteks adalah suatu diagram yang menggambarkan ruang lingkup sistem peringkas teks dalam berita bahasa indonesia. Sistem digambarkan dengan bulatan, sedangkan lingkungan diwakili oleh entitas luar yang digambarkan dengan persegi. Gambar 3.5 menggambarkan diagram konteks yang terdapat sistem peringkas teks.



Gambar 3.5 Diagram Konteks Sistem Peringkas Teks Otomatis

3.5.2. Data Flow Diagram (DFD) Level 1

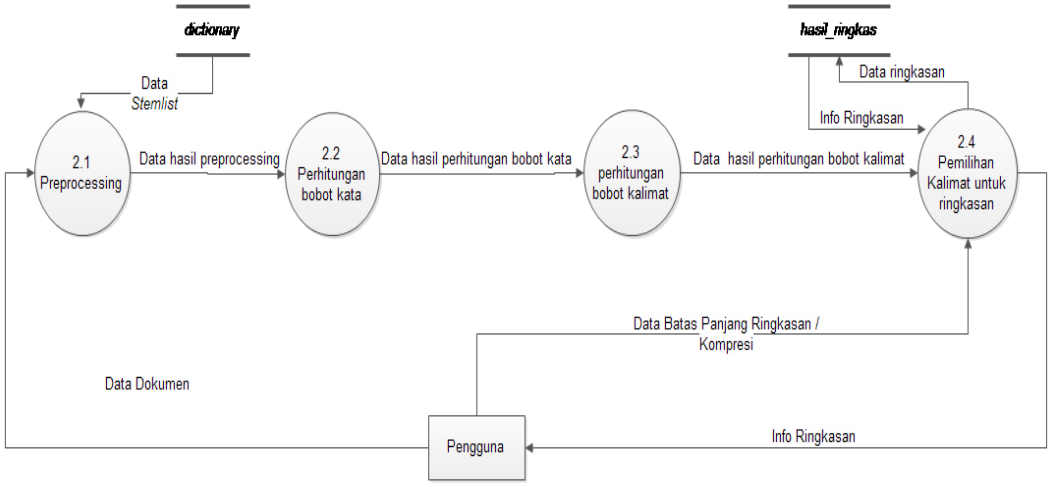
Data flow diagram level 1 berfungsi untuk menjelaskan aliran data yang terdapat di Sistem Peringkas Teks.



Gambar 3.6 Data Flow Diagram Level 1 Sistem Peringkas Teks

3.5.3. Data Flow Diagram (DFD) Level 2

Data flow diagram Level 2 berfungsi untuk menjelaskan aliran data yang terdapat di Sistem Peringkas Teks.



Gambar 3.7 Data Flow Diagram Level 1 Sistem Peringkas Teks

3.5.4. Deskripsi Proses

Deskripsi proses digunakan untuk menggambarkan proses model aliran yang terdapat pada DFD. Spesifikasi proses dari DFD yang telah dibuat dapat dijelaskan pada table deskripsi proses level 1, dan level 2.

Tabel 3.19 Spesifikasi Proses

No	Proses	Keterangan
1	No proses	1
	Nama Proses	Memilih Dokumen
	Source (Sumber)	User
	Input	Data dokumen
	Output	Info Dokumen
	Destination	User
	Logika User	1. User memilih file text data dokumen yang akan diringkas di pilih file 2. Sistem akan menampilkan dokumen yang akan diringkas 3. User dapat memilih file dokumen yang akan diringkas
2	No Proses	2
	Nama Proses	Meringkas dokumen
	Source (Sumber)	User
	Input	Data dokumen, data batas panjang ringkasan/kompresi.
	Output	Info ringkasan
	Destination	User
	Logika Proses	1. User menginputkan dokumen berupa text ke dalam sistem 2. Memilih file teks data dokumen yang akan diringkas 3. Mengatur batas panjang ringkasan / kompresi teks untuk ringkasan 4. Pilih proses untuk peringkasan teks otomatis 5. Sistem mengeluarkan data dokumen asli lalu menghasil teks ringkasan dokumen
3	No Proses	2.1
	Nama Proses	Preprocessing

	Source (sumber)	User
	Input	Data dokumen
	Output	Data hasil preprocessing
	Destination	User
	Logika Proses	1. User memilih proses untuk melihat hasil ringkasan 2. Data dokumen masuk ke tahapan praproses 3. Sistem melakukan *Preprocessing* seperti pemecahan kalimat, *case folding*, *tokenizing*, *stop word removal* dan *stemming.*
4	No Proses	2.2
	Nama Proses	Perhitungan Bobot Kata
	Source (sumber)	User
	Input	Data hasil preprocessing
	Output	Data hasil perhitungan bobot kata
	Destination	User
	Logika Proses	1. User memilih proses untuk melihat hasil ringkasan 2. Data hasil preprocessing masuk ke perhitungan bobot kata 3. Sistem melakukan perhitungan bobot kata
5	No Proses	2.3
	Nama Proses	Perhitungan Bobot Kalimat
	Source (sumber)	User
	Input	Data hasil perhitungan bobot kata
	Output	Data hasil perhitungan bobot kalimat
	Destination	User
	Logika Proses	1. User memilih proses untuk melihat hasil ringkasan 2. Data hasil perhitungan bobot kata masuk ke perhitungan bobot kalimat 3. Sistem melakukan perhitungan bobot kalimat.
6	No Proses	2.4
	Nama Proses	Pemilihan kalimat untuk ringkasan
	Source (sumber)	User
	Input	Data hasil perhitungan bobot kalimat, Data batas panjang ringkasan / kompresi
	Output	Info ringkasan

	Destination	User
	Logika Proses	1. User memilih proses dan mengatur batas panjang ringkasan/kompresi untuk melihat hasil ringkasan 2. Data hasil perhitungan bobot kalimat masuk ke pembentukan ringkasan 3. Sistem melakukan pemilihan kalimat ringkasan 4. Sistem menampilkan hasil ringkasan

3.5.5. Kamus Data

Kamus data merupakan deskripsi formal dari seluruh elemen atau aliran data yang terdapat dalam data flow diagram. Kamus data dari data flow diagram sistem peringkas teks adalah sebagai berikut :

Tabel 3.20 Data Dokumen

Nama	Data Dokumen
Where used	Memilih dokumen (1)
Deskripsi	Data yang digunakan ketika proses mencari dokumen adalah data berupa teks dengan file format txt yang digunakan dalam proses peringkasan
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

Tabel 3.21 Info Dokumen

Nama	Info Dokumen
Where used	Memilih dokumen (1)
Deskripsi	Data yang diproses ketika mencari dokumen adalah data berupa teks dengan file format txt yang digunakan dalam proses peringkasan
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

Tabel 3.22 Data Batas Panjang Ringkasan / kompresi

Nama	Data batas panjang ringkasan/kompresi
Where used	Meringkas Dokumen (2)

Deskripsi	Data yang digunakan ketika proses meringkas dokumen adalah data berupa persentase batas panjang ringkasan/kompresi.
Struktur Data	Persen kompresi
Penjelasan per struktur	Persen kompresi[0..9]

Tabel 3.23 Data Ringkasan

Nama	Data Ringkasan
Where used	Pemilihan kalimat untuk ringkasan(2.4)
Deskripsi	Data yang digunakan untuk menyimpan hasil proses peringkasan.
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

Tabel 3.24 Info Ringkasan

Nama	Info Ringkasan
Where used	Meringkas dokumen (2), *preprocessing(2.1)*, perhitungan bobot kata(2.2), perhitungan bobot kalimat(2.3), pemilihan kalimat untuk ringkasan(2.4)
Deskripsi	Data yang akan diproses adalah data inputan yang akan melalui tahapan *preprocessing*, perhitungan bobot kata, perhitungan bobot kalimat, dan menentukan kalimat untuk menampilkan hasil ringkasan
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

Tabel 3.25 Data Hasil Preprocessing

Nama	Data hasil preprocessing
Where used	preprocessing (2.1)
Deskripsi	Data yang digunakan ketika proses preprocessing adalah data dokumen.
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

Tabel 3.26 Data Hasil Perhitungan Bobot Kata

Nama	Data hasil perhitungan bobot kata
Where used	Perhitungan bobot kata

Deskripsi	Data yang digunakan ketika proses perhitungan bobot kata adalah data hasil preprocessing.
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

Tabel 3.27 Data Hasil Perhitungan Bobot Kalimat

Nama	Data hasil perhitungan bobot kalimat
Where used	Perhitungan bobot kalimat
Deskripsi	Data yang digunakan ketika proses adalah data hasil perhitungan bobot kata
Struktur Data	Teks
Penjelasan per struktur	Teks[A..Z][a..z][0..9][symbol]

3.6. Perancangan Sistem

Perancangan adalah suatu bagian dari metodologi pengembangan suatu perangkat lunak yang dilakukan setelah tahapan untuk memberikan gambaran secara terperinci. Dalam tahap ini digambarkan rancangan sistem yang akan dibangun sebelum dilakukan pengkodean ke dalam suatu bahasa pemrograman.

3.6.1. Perancangan Basis Data

Perancangan basis data pada aplikasi yang akan dibangun meliputi struktur tabel.

3.6.1.1. Struktur Tabel

Dalam perancangan basis data pada penelitian ini terdapat tabel 3.27 mengenai daftar kata dasar bahasa Indonesia.

Tabel 3.28 Struktur Tabel *Dictionary*

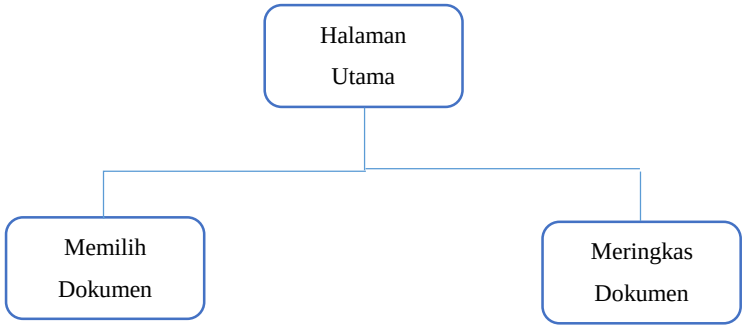
No	Atribut	Tipe Data	Keterangan
1	id_word	*Int*	*Primary key*
2	word	*Varchar[70]*	
3	Stopword	*Enum ['ya','bukan']*	

Tabel 3.29 Hasil Ringkasan

No	Atribut	Tipe Data	Keterangan
1	Id	*Int*	*Primary key*
2	Text	*Text*	

3.6.2. Perancangan Arsitektur

Setelah melakukan perancangan data pada sistem yang dibangun, maka dilakukanlah perancangan arsitektur. Perancangan arsitektur yang telah dibuat meliputi beberapa perancangan diantaranya, perancangan antarmuka, perancangan pesan, jaringan semantik dan perancangan prosedural. Berikut adalah perancangan arsitektur dapat dilihat pada gambar 3.8



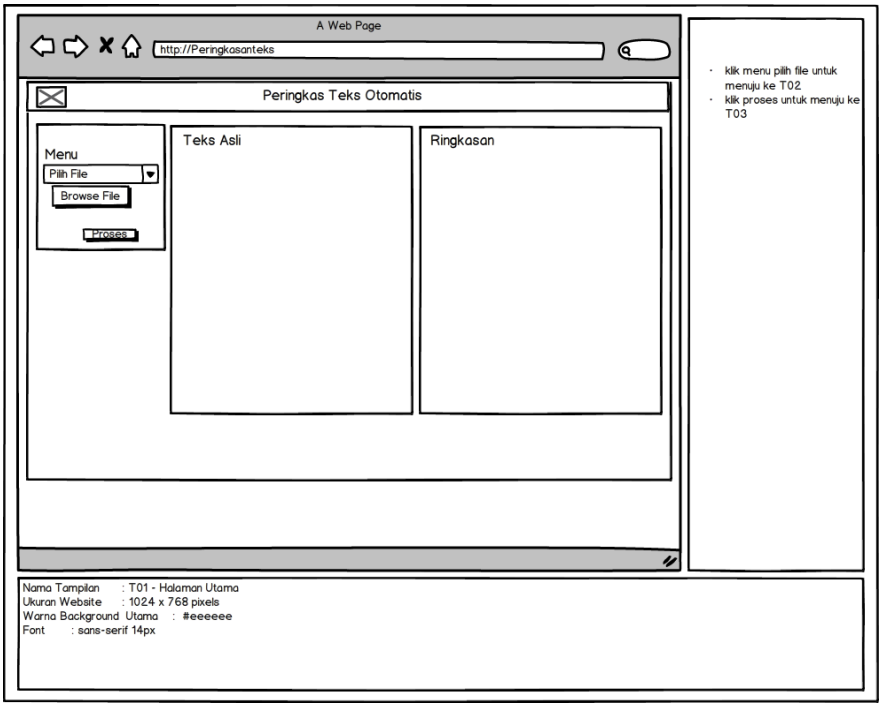
Gambar 3.8 Perancangan Arsitektur Sistem Peringkas Teks

3.6.3.Perancangan Antarmuka

Perancangan antarmuka yang telah dibuat meliputi beberapa perancangan, diantaranya perancangan antarmuka menu utama, perancangan antarmuka mencari dokumen, perancangan antarmuka meringkas dokumen, dan perancangan antarmuka klasifikasi dokumen.

3.6.4. Antarmuka Halaman Utama

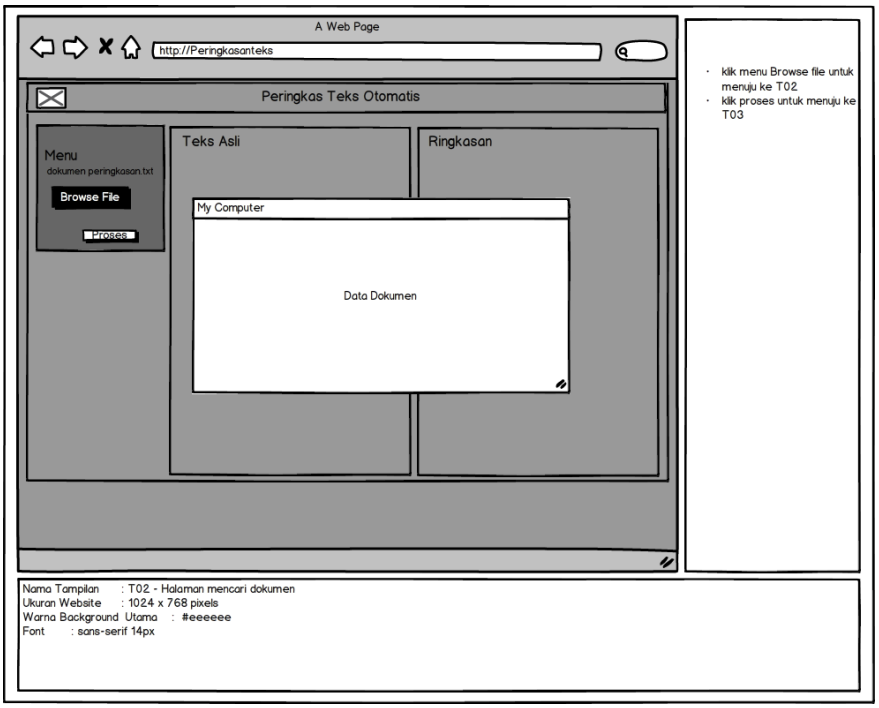
Perancangan antarmuka halaman utama dari sistem peringkas teks dapat dilihat pada gambar 3.9.



Gambar 3.9 Halaman Utama

3.6.5. Antarmuka Halaman Mencari/Pilih Dokumen

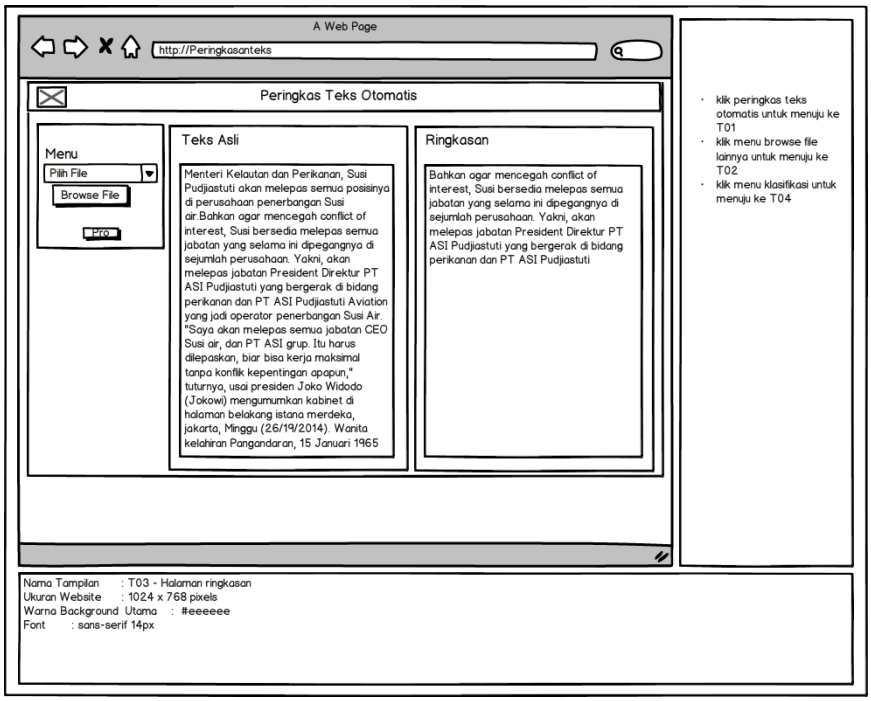
Perancangan antarmuka halaman mencari/pilih dokumen dari sistem peringkas teks dapat dilihat pada gambar 3.10



Gambar 3.10 Halaman Mencari/pilih Dokumen

3.6.6. Antarmuka Halaman Hasil Ringkasan

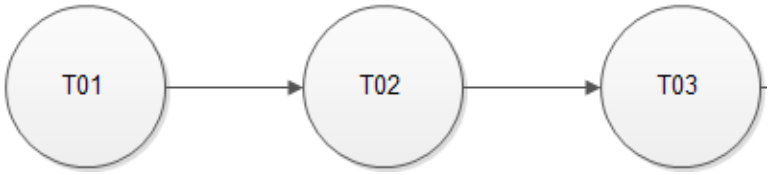
Perancangan antarmuka halaman hasil ringkasan dari sistem peringkas teks dapat dilihat pada gambar 3.11



Gambar 3.11 Halaman Hasil Peringkasan

3.6.7. Jaringan Semantik

Jaringan semantik menggambarkan keterhubungan navigasi menu dari suatu form ke form yang lainnya.jaringan semantik sistem peringkas teks dalam bahasa Indonesia dapat dilihat pada gambar 3.12.



Gambar 3.12 Jaringan Semantik