BAB III

PEMBAHASAN

3.1 Analisis Masalah

Pada bagian ini akan dilakukan analisis terhadap Information Retrieval System (IRS) dengan menggunakan Model Ruang Vektor. Hasilnya digunakan untuk mengidentifikasi dan mengevaluasi hasil implementasi dan pengujian yang akan dibahas pada bab berikutnya. Dari proses analisis akan dapat dihasilkan berbagai macam saran perbaikan terhadap sistem yang dapat dijadikan dasar dalam perancangan Information Retrieval System.

3.2 Analisis Sistem

Dari hasil survey IDC (Internasional Data Corporation) dapat diketahui bahwa dokumen yang terdapat dalam dunia maya terus bertambah. Hal ini menimbulkan kesulitan dalam menemubalikkan dokumen-dokumen tersebut. Faktor utama dalam proses pencarian suatu dokumen di dunia maya adalah bagaimana dokumen tersebut ditemubalikkan sesuai atau relevan dengan keyword atau query yang di inputkan oleh user yang akan mencari dokumen berita tersebut.

Oleh karena itu, berbagai macam penelitian di area ini telah banyak dilakukan, salah satunya adalah melalui disiplin ilmu IR (Information Retrieval).IR merupakan bidang yang mengkaji metode-metode di dalam pencarian dokumen berdasarkan representasi kebutuhan informasi berupa kata kunci,yaitu keyword atau query.

3.3 Masukan Perangkat Lunak

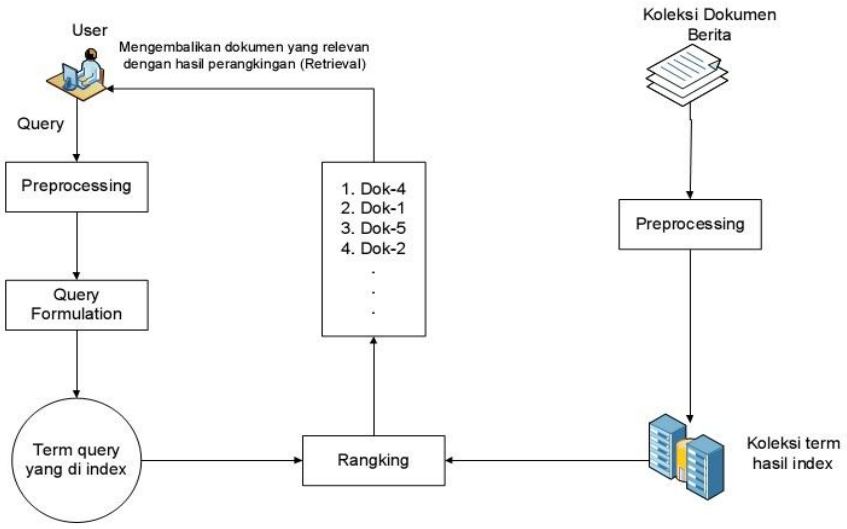
Input dari suatu sistem temu balik informasi ini adalah query dari pengguna dan koleksi dokumen. Koleksi dokumen yang digunakan dalam sistem temu balik informasi ini yaitu content berita IT berbahasa indonesia yang dapat dikumpulkan secara manual dari situs penyedia berita di internet.

3.4 Proses Yang Terjadi Didalam Perangkat Lunak

Perangkat lunak dalam laporan kerja praktek ini merupakan aplikasi Information Retrieval System dengan menggunakan Model Ruang Vektor dan metode pembobotan Term Frequency–Inverse Document Frequency (TF_IDF). Information Retrieval System ini secara umum terdapat 2 tahapan besar, yaitu preprocessing/indexing dan retrieval. Pada tahapan preprocessing mencakup penghapusan tag, tokenization, filtration, stemming, dan weigthing.Kemudian dilanjutkan dengan menyimpan setiap term yang penting ke dalam suatu index (dalam sistem ini menggunakan suatu tabel pada database MySQL).

Information Retrieval System ini akan menerima query dari pengguna kemudian memproses query tersebut dan sistem akan melakukan perhitungan kemiripan antara query dengan daftar berita yang

tersedia, untuk menghasilkan kumpulan dokumen relevan dan terurut berdasarkan ranking kerelevanannya dengan query dari pengguna.Berikut ini adalah gambar dari deskripsi umum sistem :



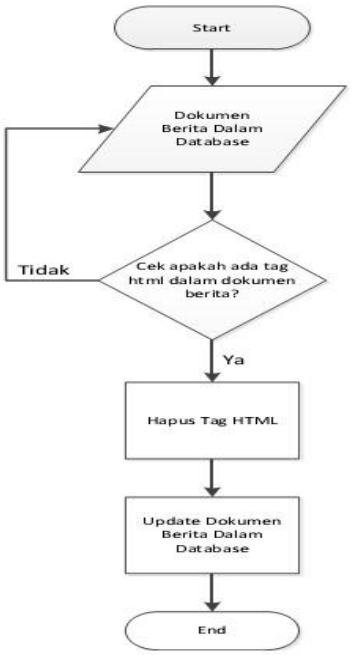
Gambar 3.1 Proses Yang Terjadi Dalam Sistem

Dalam preprocessing dilakukan beberapa tahapan untuk melakukan pengindeksan dokumen berita diantaranya penghapusan tag html,tokenization, filtration, stemming dan weighting.Berikut ini adalah penjelasan tahapan-tahapan preprocessing yang akan diterapkan dalam sistem :

3.4.1 Hapus Tag HTML

Tahap ini menghapus semua tag markup dan format khusus dari dokumen,terutama pada dokumen yang mempunyai banyak tag dan format seperti dokumen (X)HTML.Untuk penghapusan tag HTML menggunakan fungsi PHP strip_tags().Pseudocode : strip_tags(“google”);

Dibawah ini adalah Flowchart alur hapus tag HTML :



Gambar 3.2 Flowchart Hapus Tag HTML

Dokumen berita :

Dua layanan populer milik **Google**, *Picasa* dan *Blogger* tidak lama lagi akan lenyap. Bukan karena **Google** menghapusnya, tetapi karena raksasa internet itu ingin 'membuang' kedua nama produk itu untuk diganti namanya.

Sebelum tag html dihapus :

Dua layanan populer milik Google, Picasadan Blogger tidak lama lagi akan lenyap. Bukan karena Google menghapusnya, tetapi karena raksasa internet itu ingin membuang' kedua nama produk itu untuk diganti namanya.

Setelah tag html dihapus :

Dua layanan populer milik Google, Picasa dan Blogger tidak lama lagi akan lenyap. Bukan karena Google menghapusnya, tetapi karena raksasa internet itu ingin 'membuang' kedua nama produk itu untuk diganti namanya.

Dari contoh diatas bisa diketahui ada 4 kata yang mengandung tag html yang telah dihapus markup tag htmlnya.

3.4.1 Tokenization

Tokenization adalah tugas memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal atau termmed word. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua token ke bentuk huruf kecil (lower case).Dibawah ini adalah

Flowchart alur Tokenization :



Gambar 3.3 Flowchart Tokenization

Untuk pemisahan kalimat ke bentuk kata menggunakan fungsi PHP explode().

Pseudocode :

explode(“ “, “Dua layanan populer milik Google”);

Contoh tokenization :

Input :

Dua layanan populer milik Google, Picasa dan Blogger tidak lama lagi akan lenyap. Bukan karena Google menghapusnya, tetapi karena raksasa internet itu ingin 'membuang' kedua nama produk itu untuk diganti namanya.



3.4.1 Filtration

Filtration adalah proses penghapusan kata buang yaitu kata sambung, kata depan, kata ganti, dll. Contoh stop words dalam bahasa Indonesia : yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dan, tersebut, pada, dengan, adalah, yaitu, ke, tak, tidak, di,pada, jika, maka, ada, pun, lain, saja, hanya, namun, seperti, kemudian, karena, untuk, dll. Proses filtration menggunakan fungsi PHP str_replace().

Dibawah ini adalah array stopword yang telah di inputkan beserta contoh penggunaan fungsi str_replace() :

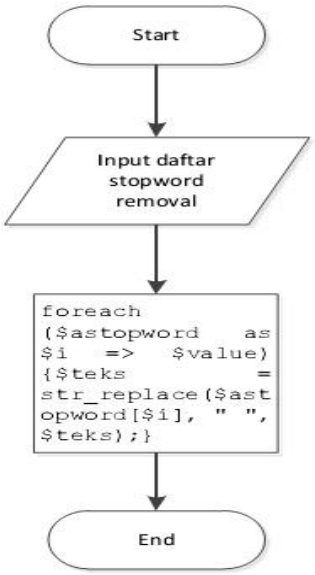
$astopword = array (" yang ", " juga ", " dari ", " dia ", " kami ", " kamu ", "aku", " ini ", " itu ", " atau ", " dan ", " tersebut ", " pada ", " dengan ", " adalah ", " yaitu ", " ke ", " tak "," tidak "," di ", " pada "," jika "," maka "," ada ", " pun "," lain "," saja ", " hanya "," namun "," seperti "," kemudian "," untuk "," sudah ", " belum ", " memang "," sama ", " apa "," apakah "," oleh ", " sangat "," sehingga "," memang "," dimana "," sedangkan ", " selain "," tapi "," tetapi "," tanpa "," tiada "," dengan "," lain "," akan "," lainnya "," tiada "," para ", " saat "," sebelum "," sesudah "," karena ");

Data array akan dibaca oleh fungsi foreach :

```
foreach ($astopword as $i => $value) {
$teks = str_replace($astopword[$i], " ", $teks);
}
```

$teks adalah dokumen berita yana akan diproses oleh filtration.

Dibawah ini flowchart proses filtration :



Contoh Sebelum *filtration* :

dua layanan populer milik google picasa dan blogger tidak lama lagi akan lenyap bukan karena google menghapusnya tetapi karena raksasa internet itu ingin membuang kedua nama produk itu untuk diganti namanya

Setelah *Filtration* :

dua layanan populer milik google picasa blogger lama lenyap google menghapus raksasa internet ingin membuang kedua nama produk diganti nama

3.4.4 Stemming

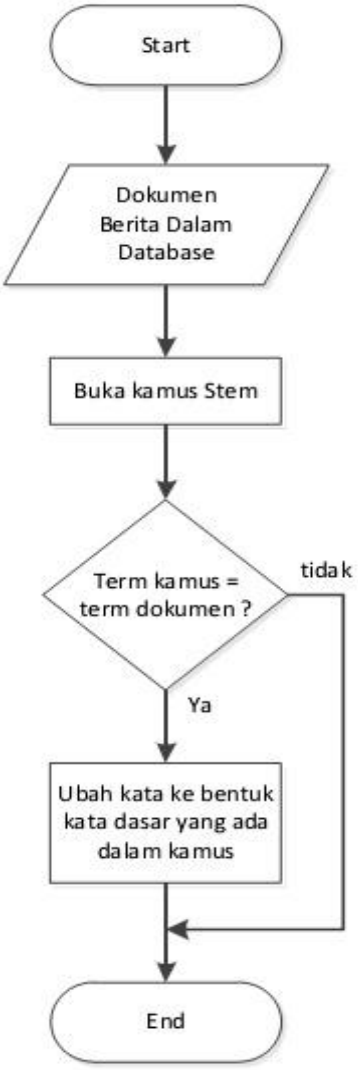
Stemming adalah pengubahan kata ke bentuk kata dasar atau penghapusan imbuhan. Stemming disini menggunakan kamus daftar kata berimbuhan yang mempunyai kata dasarnya dengan cara membandingkan kata-kata yang ada dalam dokumen berita dengan daftar kamus stem. Proses stemming menggunakan fungsi PHP str_replace. Berikut ini adalah contoh penggunaanya :

```
//buka tabel tbstem dan bandingkan dengan berita

$restem = mysql_query("SELECT * FROM stem ORDER BY id");

while($rowstem = mysql_fetch_array($restem)) {
        $teks=str_replace($rowstem['term'],$rowstem['stem'],
$teks);
 }
```

Dibawah ini *flowchart* proses *stemming* :



Contoh kamus stemming

No	Term	Stem
1	layanan	Layan
2	menghapus	Hapus
3	membuang	Buang
4	Kedua	Dua
5	Diganti	Ganti
6	namanya	Nama

Contoh stemming

dua layanan populer milik google picasa blogger lama lenyap google menghapus raksasa internet ingin membuang kedua nama produk diganti nama

Term yang berubah saat proses steming

Mengubah → ubah , Tertentu → tentu

Tahapan → *tahap, Dilaporkan* → *lapor*

Sebelum → *belum,*

Setelah Stemming

dua layan popular milik google picasa blogger lama lenyap google hapus raksasa internet ingin buang dua nama produk ganti nama

3.4.5 Weighting

Pembobotan kata sangat berpengaruh dalam menentukan kemiripan antara dokumen dengan query. Apabila bobot tiap kata dapat ditentukan dengan tepat, diharapkan hasil perhitungan kemiripan teks akan menghasilkan perangkingan dokumen yang baik.

Bobot term di dalam Information Retrieval System (W) dihitung menggunakan tf-idf yang didefinisikan sebagai berikut.

W= tf ×idf

Diketahui :

W = Bobot term

tf = banyaknya kemunculan suatu kata dalam suatu dokumen/berita.

Idf = $log\ (N/dji)$

Diketahui :

N = Jumlah artikel dalam koleksi dokumen

dji = Banyaknya Dokumen (j) yang mengandu term (i)

Log2 = digunakan untuk memperkecil pengaruhnya relatif terhadap tfij

Contoh kasus

Berita 1 : Saat mengubah suatu proses kita harus mengetahui tahapan yang

Berita 2 : Saat kita telah mendegar banyak proses berita yang merugikan

Berita 3 : Tokenization adalah tugas proses memisahkan deretan kata didalam kalimat,

Diketahui :

Berita = 3 (Jumlah berita)

Idf = $log\ (N/dji)$

dji = Banyaknya berita (j) yang mengandu term (i)

Tf Berita 1 = Banyaknya muncul kata di Berita 1

Tf Berita 2 = Banyaknya muncul kata di Berita 2

Tf Berita 3 = Banyaknya muncul kata di Berita 3

W Berita 1 = Bobot kata di Berita 1

W Berita 2 = Bobot kata di Berita 2

W Berita 3 = Bobot kata di Berita 3

Tabel 3.2 Pembobotan Kata

Kata	Tf D_1	Tf D_2	Tf D_3	d_{ji}	$\frac{N}{d_{ji}}$	Idf	W D_1	W D_2	W D_3
Dua	1	0	0	1	3/1 = 3	0.48	0.48	0	0
Layan	1	0	0	1	3/1 = 3	0.48	0.48	0	0
Popular	1	0	0	1	3/1 = 3	0.48	0.48	0	0
Milik	1	1	0	2	3/2 = 1.5	0.18	0.18	0.18	0
Google	1	1	2	4	¾ = 0.75	0.12	0.12	0.12	0.24
Picasa	1	1	1	3	3/3 = 1	0	0	0	0
Blogger	1	1	0	2	3 / 2 =1.5	0.18	0.18	0.18	0
Lama	1	0	0	1	3/1 =3	0.48	0.48	0	0
Lenyap	1	0	0	1	3/1 = 3	0.48	0.48	0	0
Nama	0	0	1	1	3/1 = 3	0.48	0	0	0.48
Ubah	0	0	1	1	3/1 = 3	0.48	0	0	0.48
Dalam	0	0	1	1	3/1 = 3	0.48	0	0	0.48
Cari	0	0	1	1	3/1 = 3	0.48	0	0	0.48

3.5 Keluaran Perangkat Lunak

Output yang diharapkan dari Information Retrieval System ini adalah dokumen yang dianggap relevan oleh sistem berupa berita yang telah terurut berdasarkan relevansi dengan query dari pengguna.

3.6 Analisis dan Kebutuhan Non Fungsional

Analisis dan kebutuhan non fungsional meliputi analisis dan kebutuhan perangkat keras serta analisis dan kebutuhan perangkat lunak.

3.6.1 Analisis dan Kebutuhan Perangkat Lunak

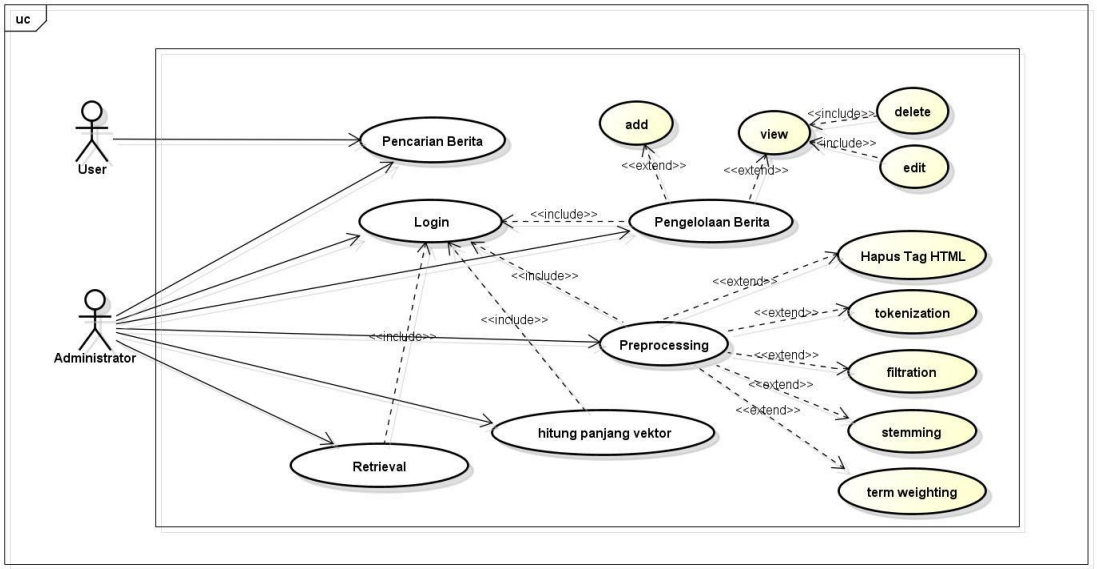
Spesifikasi perangkat lunak yang digunakan dalam Information Retrieval System ini adalah :

- Sistem Operasi Microsoft Windows 7 Professional
- Software wampp sebagai Web Server
- PHP sebagai bahasa pemrograman web
- Mysql sebagai sistem manajemen basis data
- Software utility Dreamweaver CS5.

3.6.2 Use Case Diagram

Pemodelan use case adalah pemodelan sistem dari perspektif pandangan pemakai akhir (end user). Model use case adalah pandangan dari luar sistem, sementara model rancangan adalah pandangan dari dalam. Model use case menangkap penggunaan penggunaan sistem, sedangkan model rancangan merepresentasikan pembangunan dari sistem.

Dibawah ini adalah Use Case Diagram Information Retrieval :



3.7 Identifikasi Aktor

Terdapat dua aktor yang berhubungan dengan Information Retrieval System.Aktor dapat diidentifikasikan sebagai berikut :

a. Aktor pertama ialah Administrator, mempunyai hak akses untuk pengelolaan dokumen berita, melakukan indexing, dan menerapkan model ruang vektor pada dokumen berita juga mempunyai hak akses ke halaman user untuk melakukan pencarian.
b. Aktor kedua ialah User, hanya mempunyai hak akses di halaman user untuk melakukan pencarian dokumen berita.

3.7.1 Identifikasi Use Case

Tabel 3.3 Identifikasi Use Case Information Retrieval System

No.	*Use Case*	Deskripsi
1	*Login*	Proses ini melakukan autentikasi terhadap aktor yang menjalankan sistem untuk memastikan bahwa orang yang bersangkutan memiliki hak akses terhadap sistem.

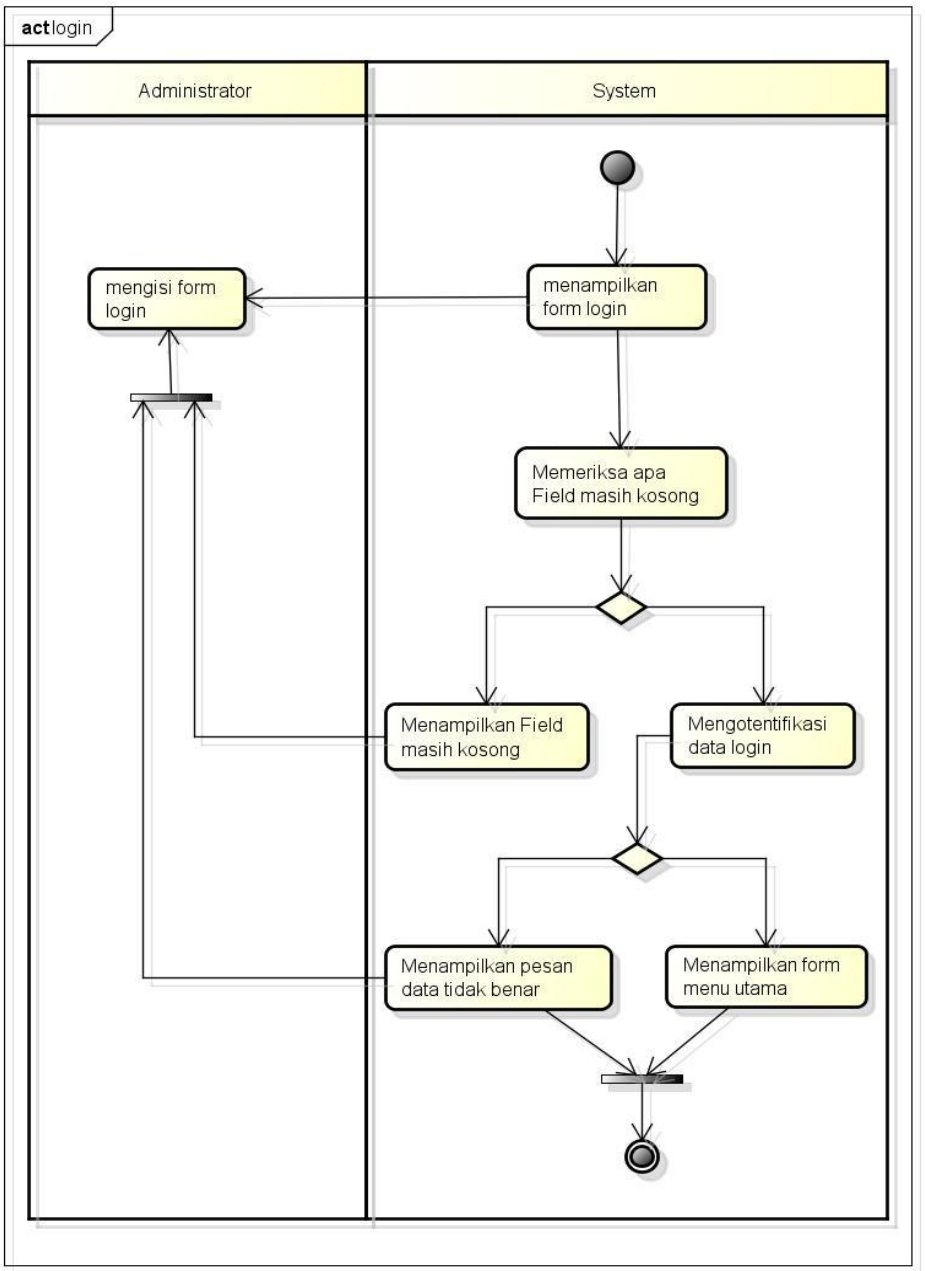
2	*Add*	Proses bagi admin untuk menambahkan dokumen berita
3	*View*	Proses bagi admin untuk melihat dokumen berita
4	*Edit*	Proses bagi admin untuk mengubah dokumen berita
5	*Delete*	Proses bagi admin untuk menghapus dokumen berita
6	Hapus *tag html*	Proses untuk menghapus *tag html*
7	*Tokenization*	Proses untuk memecah dokumen berita menjadi potongan-potongan kata
8	*Filtration*	Proses untuk *stopword* menghilangkan kata-kata yang tidak diperlukan untuk pengindeksan dan menghilangkan tanda baca.
9	*Stemming*	Proses untuk mengubah kata ke bentuk kata dasar atau penghapusan imbuhan.
10	*Term weighting*	Proses untuk pemberian bobot tf.idf terhadap term
11	Hitung panjang vektor	Proses untuk menghitung panjang vektor dokumen
12	*Retrieval*	Proses bagi admin untuk menemubalikkan dokumen berita.
13	Pencarian Berita	Proses bagi user untuk pencarian dokumen berita

3.2.1.1 *Use Case Login*

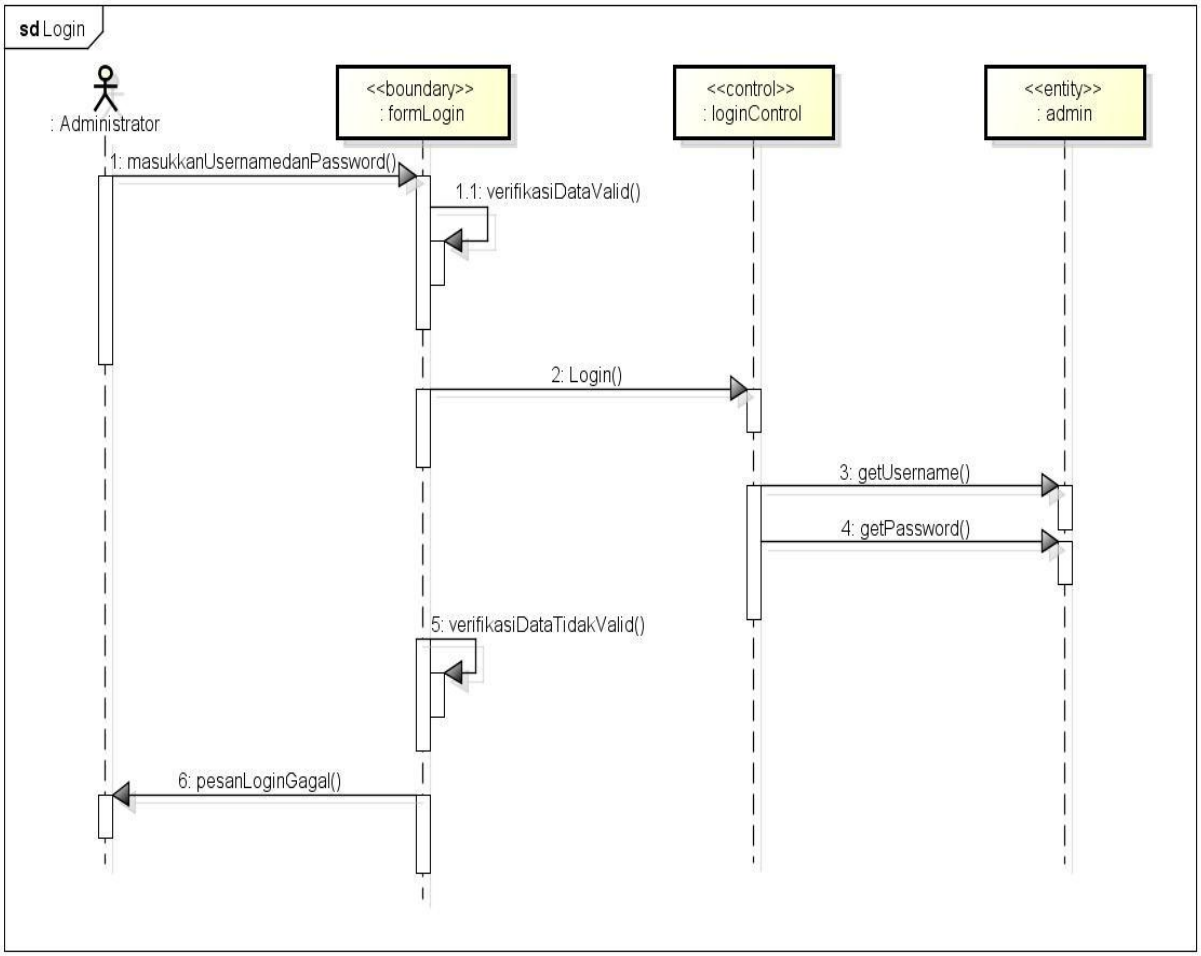
Proses ini melakukan autentikasi terhadap aktor yang menjalankan sistem untuk memastikan bahwa orang yang bersangkutan memiliki hak akses terhadap sistem. *Use case login* diterangkan dan dimodelkan dengan skenario, *activity diagram* dan *sequence diagram.*

Tabel 3.4 Skenario *Use Case* Login

Identifikasi	
Nama	*Login*
Tujuan	Masuk ke dalam sistem
Deskripsi	Proses *login* merupakan proses autentifikasi untuk menggunakan sistem.
Aktor	*Administrator*
Skenario Utama	
Kondisi awal	*Form login* ditampilkan
Aksi actor	Reaksi sistem
	1) Menampilkan *form login*
2) Mengisi *Form Login*	2) Cek *field* masih kosong
	3) Mengautentifikasi data *login* dengan data *username* pada basis data
	4) Bila cocok sistem menampilkan halaman menu utama
Skenario Alternatif (Auntentifikasi Gagal)	
Aksi Aktor	Reaksi Sistem
	1) Menampilkan Pesan bahwa data *username* tidak benar
2) Mengisi kembali *Form Login*	3) Mengautentifikasi data *username* dan *password* pada basis data
	4) Bila cocok sistem menampilkan halaman menu utama
Kondisi akhir	Menu utama ditampilkan



Gambar 3.8 *Activity Diagram* Login



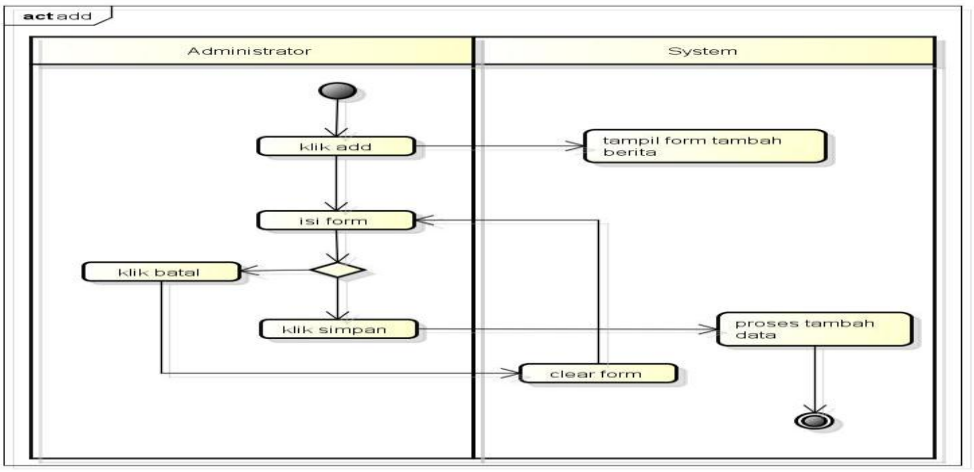
Gambar 3.9 *Sequence Diagram Login*

3.2.1.2 *Use Case Add*

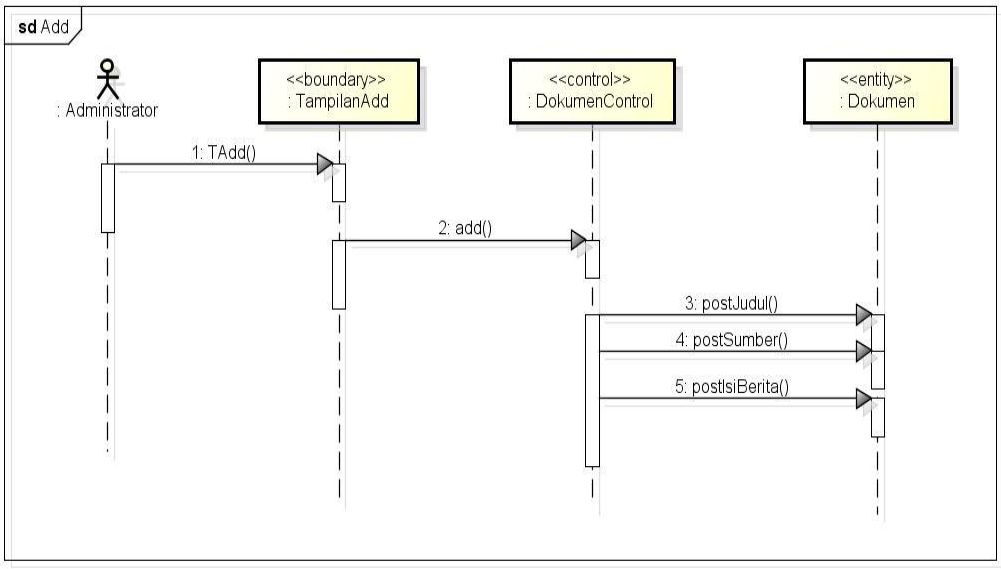
Proses ini melakukan penambahan dokumen berita. *Use case add* diterangkan dan dimodelkan dengan skenario, *activity diagram* dan *sequence diagram.*

Tabel 3.5 Skenario *Use Case Add*

Identifikasi	
Nama	*Add*
Tujuan	Menambah dokumen berita
Deskripsi	Proses *Add* merupakan proses untuk penambahan dokumen berita ke dalam *Information Retrieval System*
Aktor	*Administrator*
Skenario Utama	
Kondisi awal	Klik tombol add
Aksi aktor	Reaksi sistem
1. Klik add	2. Menampilkan *Form* tambah dokumen berita.
3. Mengisi *Form* Dokumen berita	
4. Klik tombol simpan	5. Memproses penambahan data
Skenario Alternatif (Auntentifikasi Gagal)	
Aksi Aktor	Reaksi Sistem
1. Klik tombol batal	2. Kosongkan *form*
Kondisi akhir	Tampil Sukses



Gambar 3.10 Activity Diagram Add



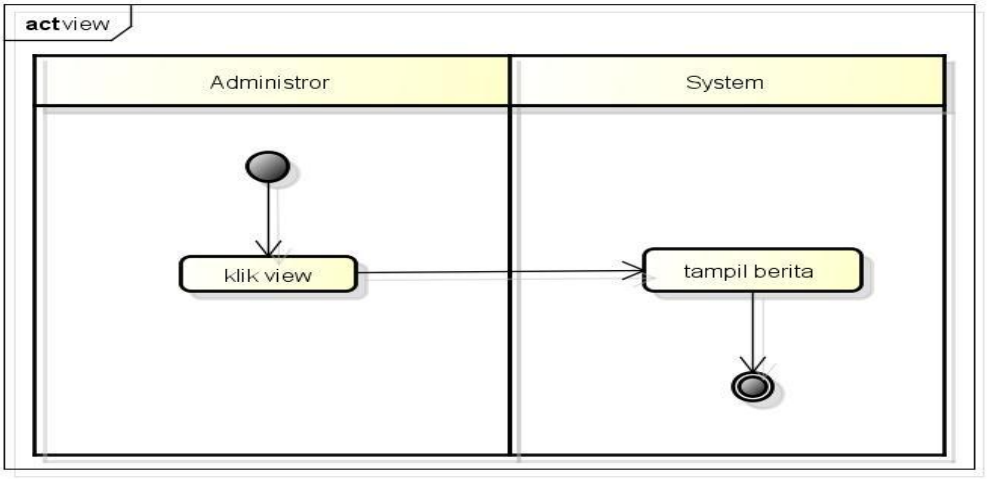
Gambar 3.11 *Sequence Diagram Add*

3.2.1.3 *Use Case View*

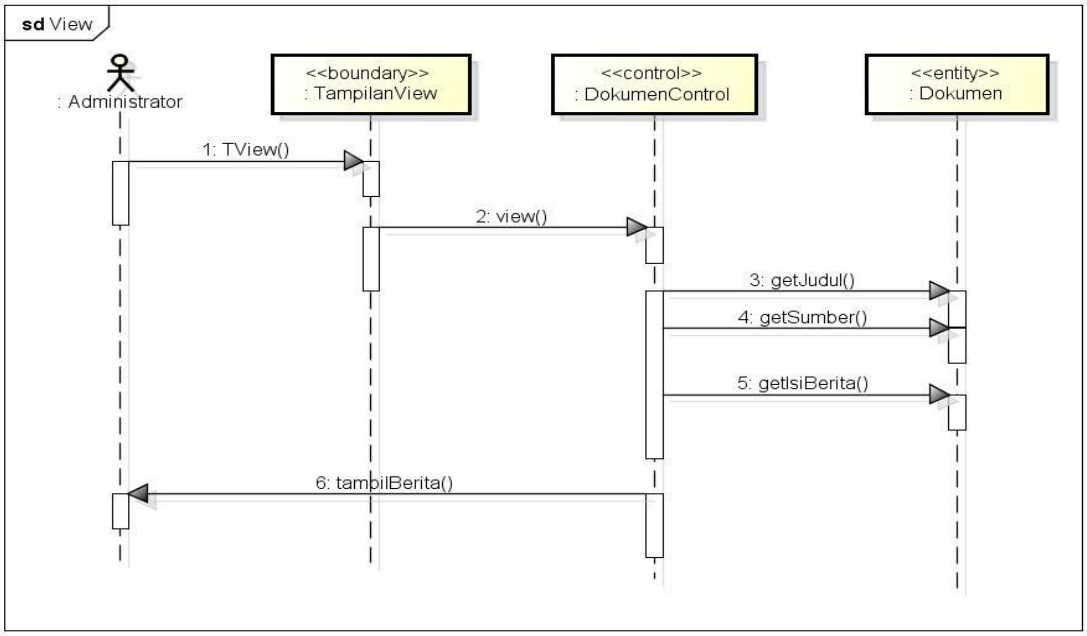
Proses ini menampilkan dokumen berita. *Use case* view diterangkan dan dimodelkan dengan skenario, *activity diagram* dan *sequence diagram.*

Tabel 3.6 Skenario *Use Case* View

Identifikasi	
Nama	*View*
Tujuan	Menampilkan dokumen berita
Deskripsi	Proses *View* untuk menampilkan berita.
Aktor	*Administrator*
Skenario Utama	
Kondisi awal	Klik tombol *view*
Aksi actor	Reaksi sistem
1. Klik view	2. Tampilkan dokumen berita
Kondisi akhir	Admin berhasil *view* berita



Gambar 3.12 ***Activity Diagram View***



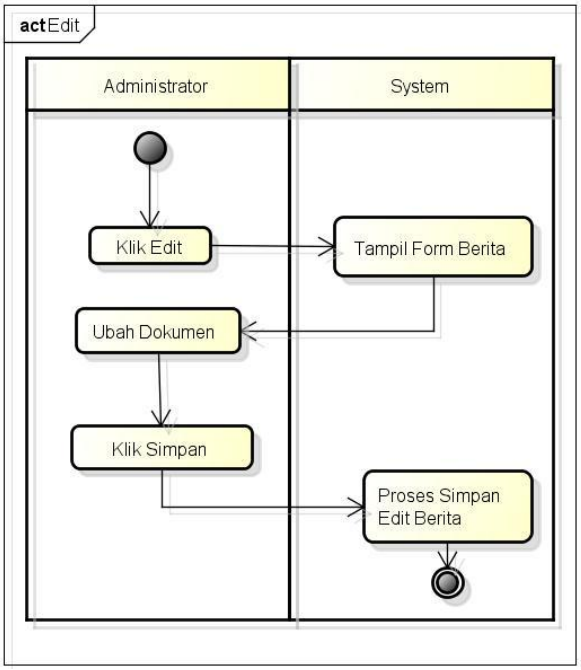
Gambar 3.13 ***Sequence Diagram View***

3.2.1.4 *Use Case Edit*

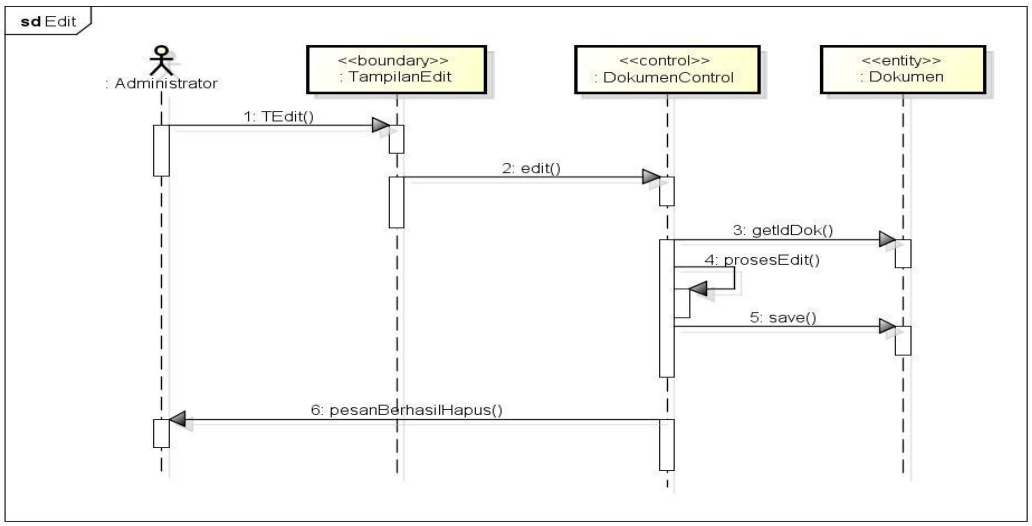
Proses ini untuk mengubah dokumen berita. *Use case edit* diterangkan dan dimodelkan dengan skenario, *activity diagram* dan *sequence diagram.*

Tabel 3.7 Skenario *Use Case Edit*

Identifikasi	
Nama	*Edit*
Tujuan	Mengubah dokumen berita
Deskripsi	Proses untuk mengubah dokumen berita
Aktor	*Administrator*
Skenario Utama	
Kondisi awal	Klik tombol *view*
Aksi actor	Reaksi sistem
1. Klik *Edit*	2. Tampilkan *form* Edit Berita
3. Ubah Dokumen Berita	
4. Klik Simpan	5. Simpan Hasil Edit
Kondisi akhir	Admin berhasil mengubah dokumen berita



Gambar 3.14 *Activity Diagram Edit*



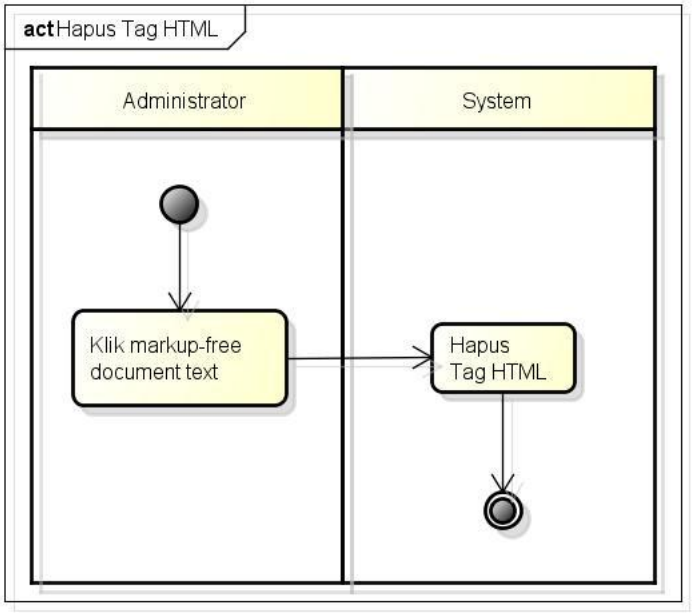
Gambar 3.17 Sequence Diagram Delete

3.2.1.5 *Use Case* Hapus *Tag HTML*

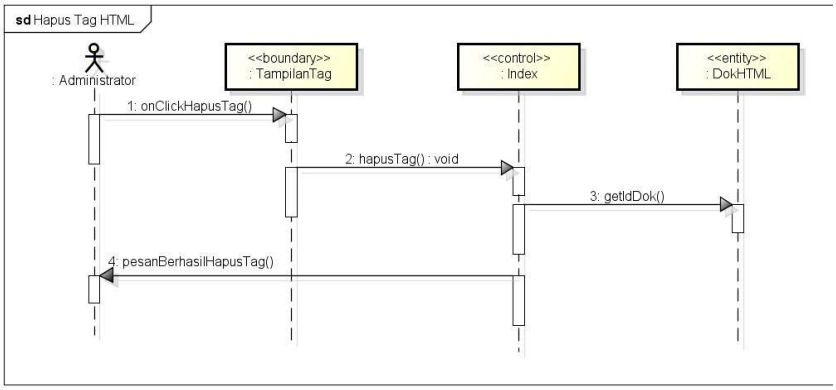
Proses untuk menghitung panjang vektor dokumen berita. *Use case* hitung panjang vector diterangkan dan dimodelkan dengan skenario, *activity diagram* dan *sequence diagram.*

Tabel 3.8 Skenario *Use Case* Hapus *Tag HTML*

Identifikasi	
Nama	Hapus *Tag HTML*
Tujuan	Menghapus *Tag HTML*
Deskripsi	Proses untuk menghapus *Tag HTML*
Aktor	*Administrator*
Skenario Utama	
Kondisi awal	Klik tombol *Markup-free document text*
Aksi actor	Reaksi sistem
1. Klik *Markup-free document text*	2. Hapus *Tag HTML*
Kondisi akhir	Admin berhasil menghapus dokumen berita



Gambar 3.18 *Activity Diagram* Hapus *Tag HTML*



Gambar 3.19 Sequence Diagram Hapus Tag HTML

3.3 Analisis dan Kebutuhan Perangkat Keras

Spesifikasi perangkat keras yang digunakan dalam Information Retrieval System ini adalah :

- Processor : AMD Turion™ X2 Dual-Core Mobile RM-74 (2 CPUs) 2.2 GHz.
- RAM : 3 Giga
- VGA : Onboard
- HDD : 320 Giga
- Monitor : 15”

3.4 Perancangan Sistem

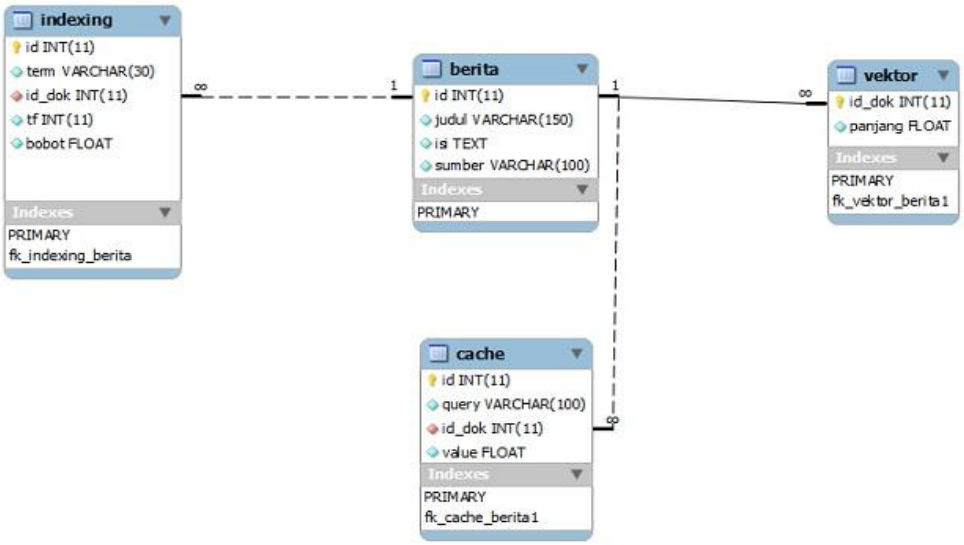
Perancangan akan dimulai setelah tahap analisis terhadap sistem selesai dilakukan. Perancangan dapat didefinisikan sebagai proses aplikasi berbagai teknik dan prinsip bagi tujuan pendefinisian suatu perangkat, suatu proses atau sistem dalam detail yang memadai untuk memungkinkan realisasi fisiknya. Perancangan digambarkan sebagai proses multi-langkah dimana representasi struktur data, struktur program, karakteristik interface, dan detail prosedur, disintesis dari persyaratan informasi.

3.5 Perancangan Basis Data

Perancangan data mentransformasi model domain informasi yang dibuat selama analisis ke dalam struktur data yang akan diperlukan untuk mengimplementasi perangkat lunak.

3.5.1 Skema Relasi

Proses relasi antar file merupakan gabungan antar file yang mempunyai kunci utama yang sama, sehingga file-file tersebut menjadi satu kesatuan yang dihubungkan oleh field kunci tersebut. Pada proses ini elemen-elemen data dikelompokkan menjadi satu file database beserta entitas dan hubungannya. Berikut adalah skema relasi antar tabel yang ada di aplikasi Information Retrieval System:



Gambar 3.4 Skema Relasi

3.5.2 Struktur Tabel

Tabel adalah sekumpulan data atau informasi spesifik tentang subjek tertentu yang disusun dalam bentuk kolom dan baris. Tabel adalah komponen utama dan pertama dari sebuah database. Struktur tabel yang terdapat dalam database Information Retrieval System sebagai berikut :

a. Tabel Berita

Tabel Berita digunakan untuk menyimpan dokumen-dokumen berita.

Tabel 3.5 Tabel Berita

Nama field	Tipe	Size	Keterangan
id	Integer	11	*Primary key*
judul	varchar	150	
isi	Text		
sumber	varchar	100	

b. Tabel Indexing

Tabel Indexing digunakan untuk menyimpan data hasil preprocessing dan weighting (pembobotan).

Tabel 3.6 Tabel Indexing

Nama field	Tipe	Size	Keterangan
id	Integer	11	*primary key*
term	Varchar	30	
id_dok	Integer	11	*foreign key, reference* berita(id)
tf	Integer	11	
bobot	Float		

c. Tabel Stem

Tabel Stem digunakan untuk menyimpan kamus atau koleksi term-term yang akan diubah kebentuk kata dasarnya.

Tabel 3.7 Tabel Stem

Nama field	Tipe	Size	Keterangan
id	Integer	11	primary key
term	Varchar	50	
Stem	Varchar	50	

d. Tabel Vektor

Tabel Vektor digunakan untuk menyimpan hasil perhitungan panjang vektor dari tiap dokumen berita.

Tabel 3.8 Tabel Vektor

Nama field	Tipe	Size	Keterangan
id_dok	Integer	11	foreign key, reference berita(id)
Panjang	Float		

e. Tabel Cache

Tabel Cache digunakan untuk menyimpan history query atau keyword dan untuk menyimpan nilai kemiripan antara query dan dokumen berita.

Tabel 3.9 Tabel Cache

Nama field	Tipe	Size	Keterangan
Id	Integer	11	*primary key*
Query	Varchar	100	
id_dok	Integer	11	*foreign key, reference* berita(id)
Value	Float		

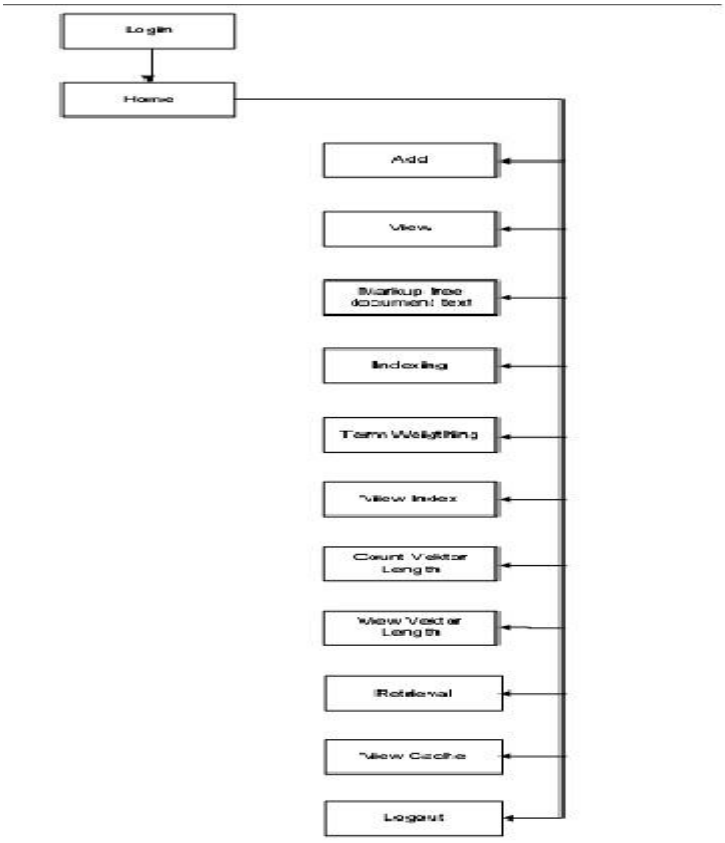
f. Tabel Admin

Tabel Admin digunakan untuk menyimpan data login administrator.

Tabel 3.10 Tabel Admin

Nama field	Tipe	Size	Keterangan
username	Varchar	20	*primary key*
password	Varchar	50	

3.6 Perancangan Struktur Menu

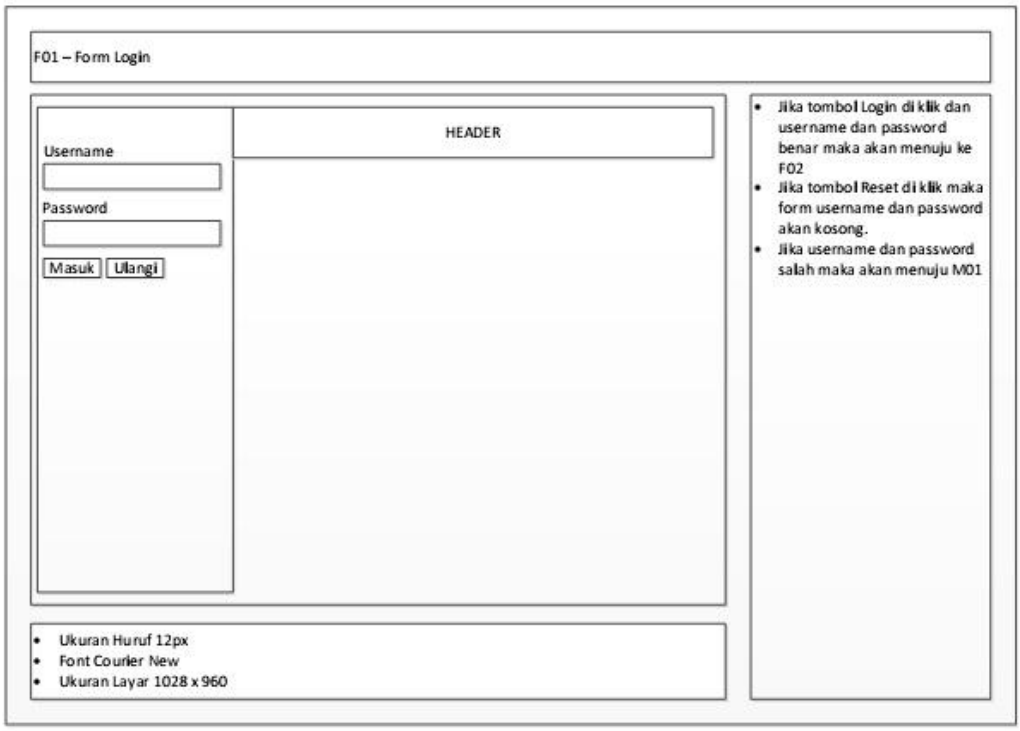


Gambar 3.36 Struktur Menu

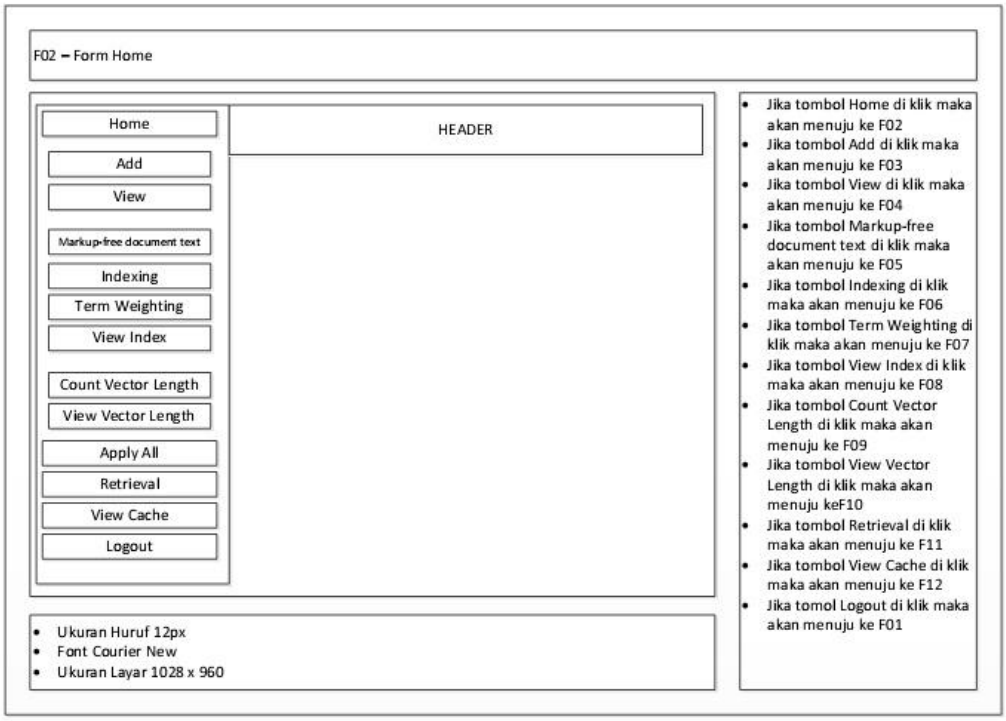
3.7 Perancangan Antar Muka

Perancangan antarmuka (interface) menggambarkan bagaimana perangkat lunak berkomunikasi dengan dirinya sendiri, dengan sistem yang berinteroperasi dengannnya,dan dengan manusia yang menggunakannya.

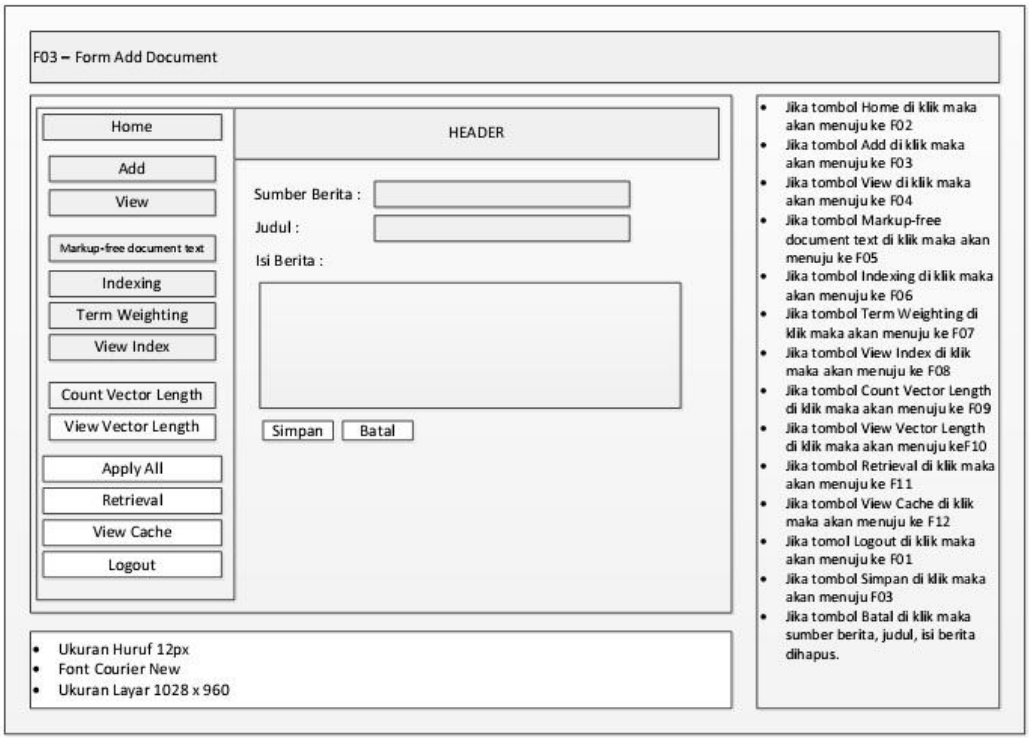
3.8 Perancangan Tampilan



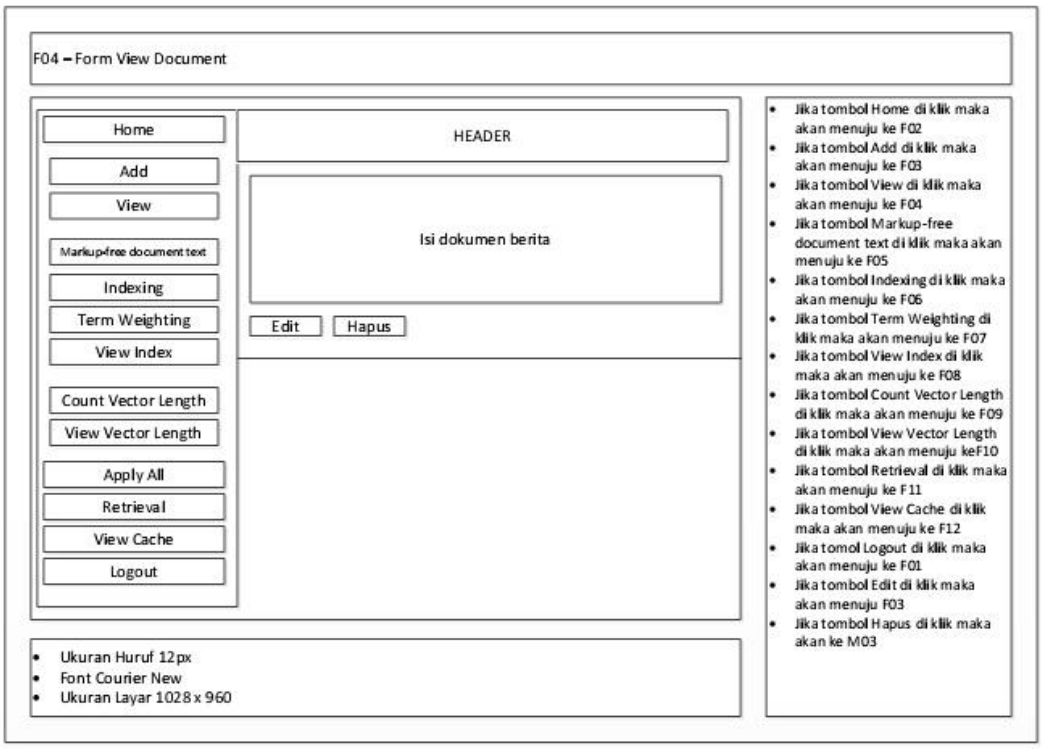
Gambar 3.37 Tampilan Halaman Login



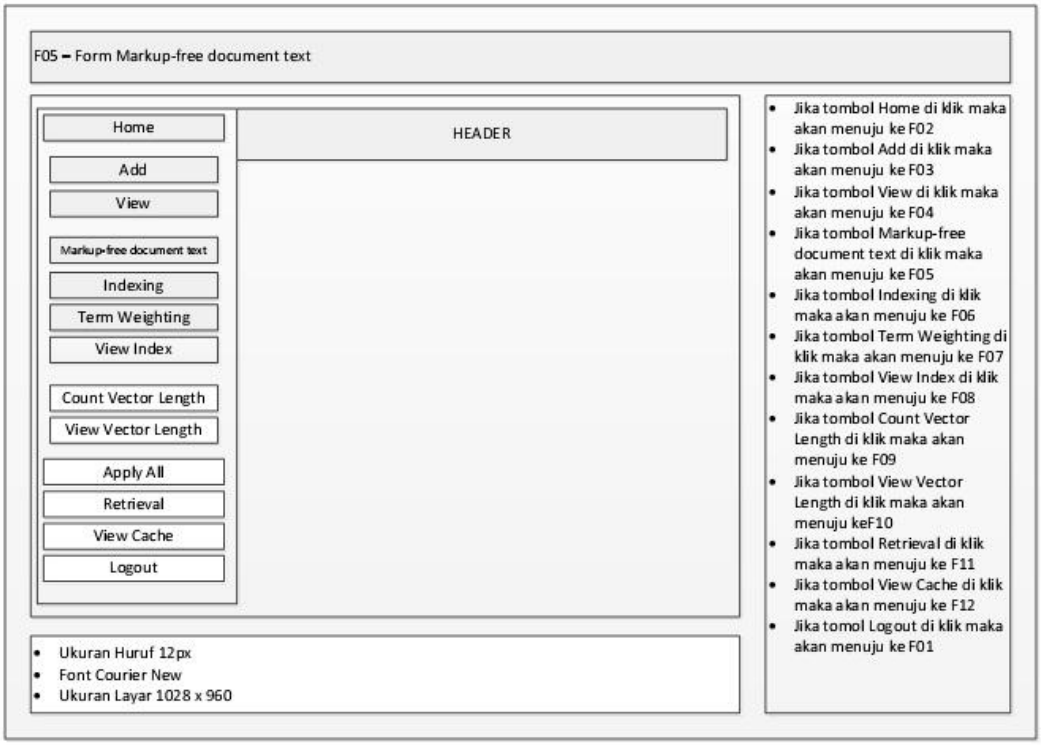
Gambar 3.38 Tampilan Halaman Home



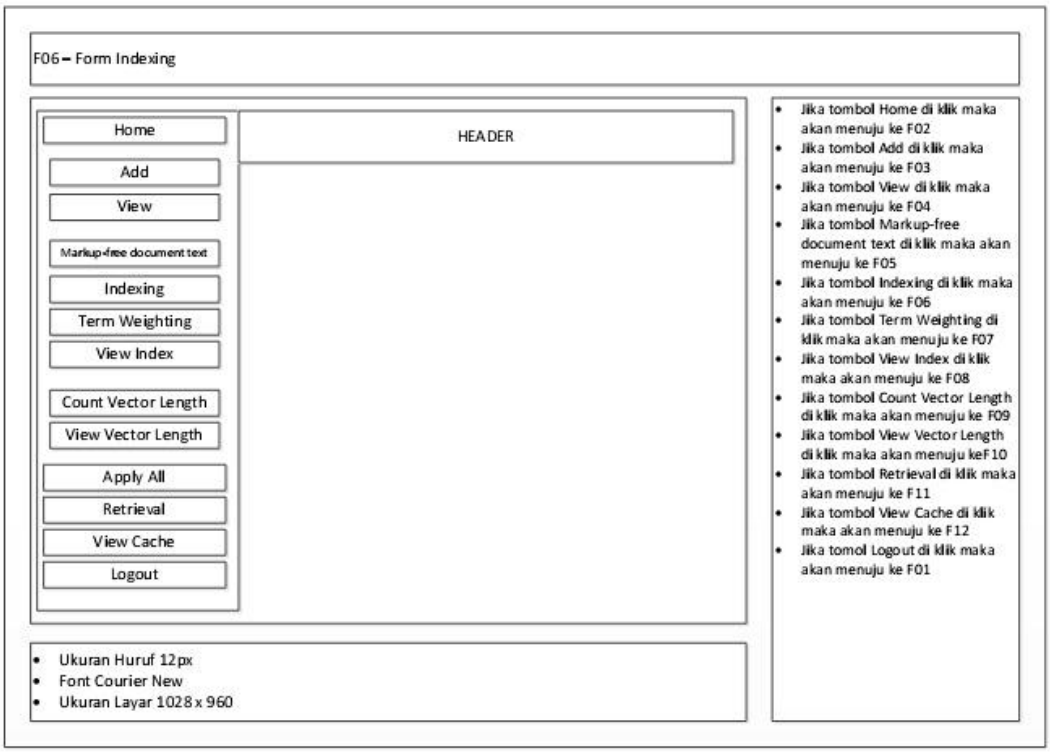
Gambar 3.39 Tampilan Halaman Tambah Dokumen Berita



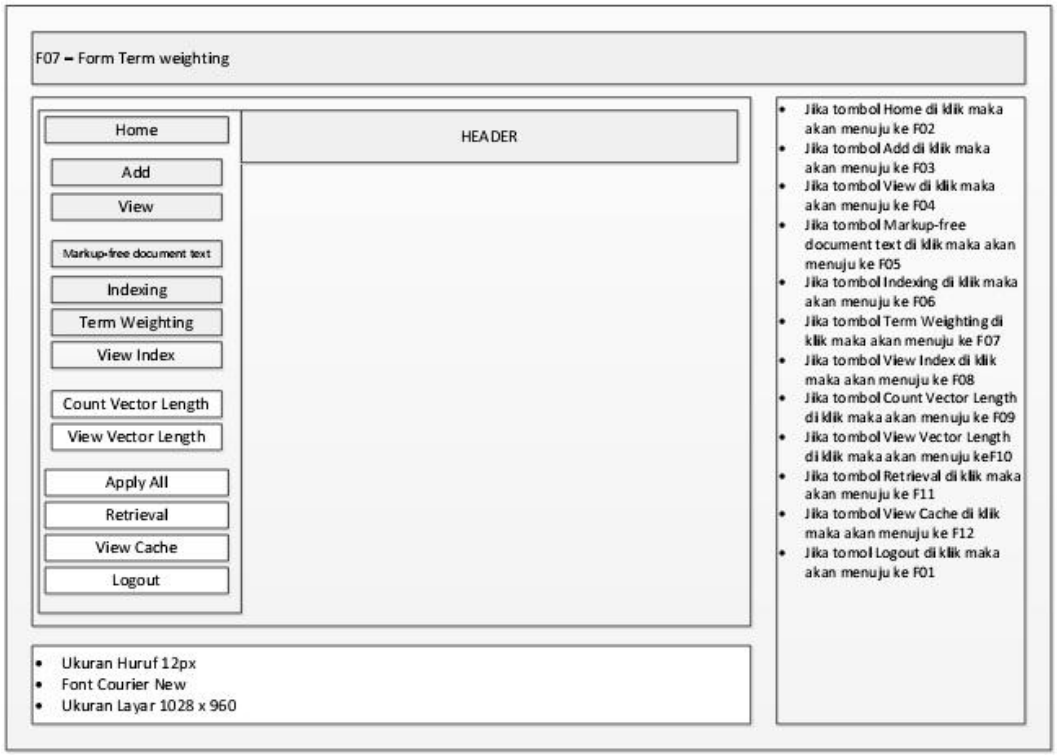
Gambar 3.40 Tampilan Halaman View Document



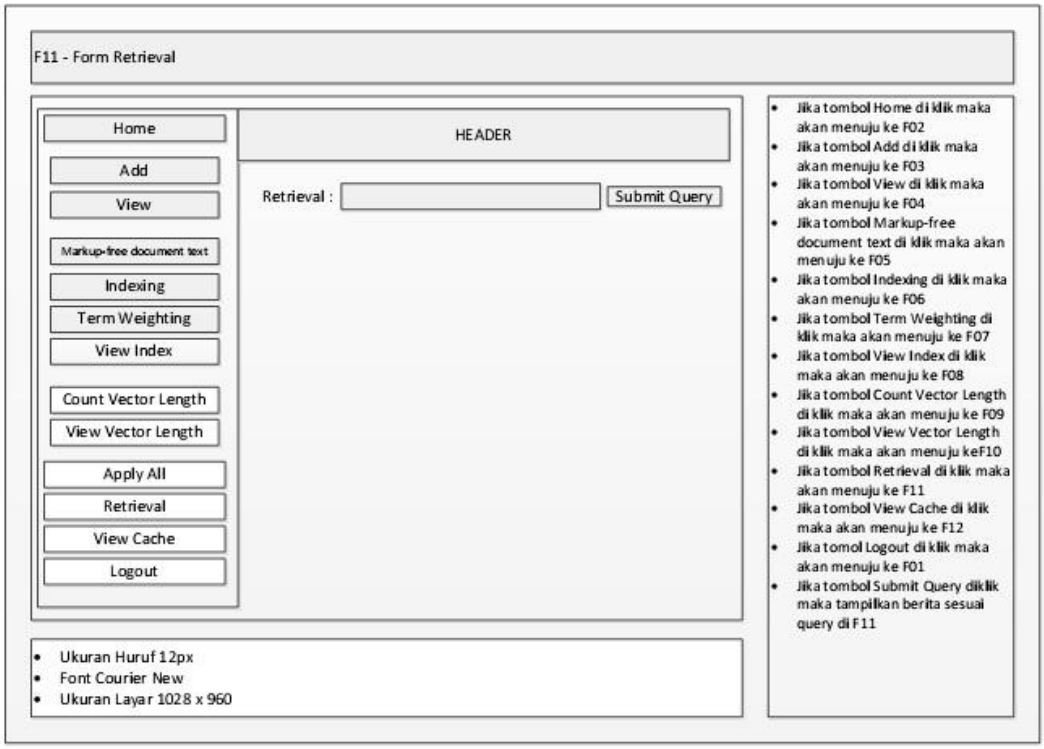
Gambar 3.41 Halaman Hapus Tag HTML



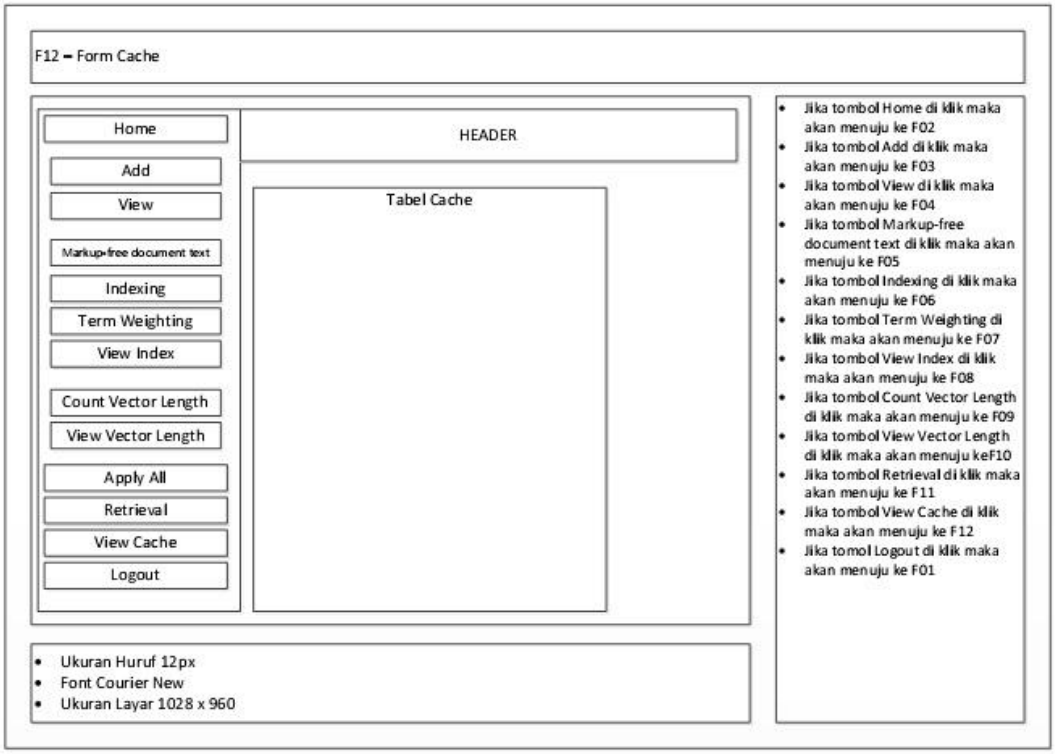
Gambar 3.42 Tampilan Halaman Indexing



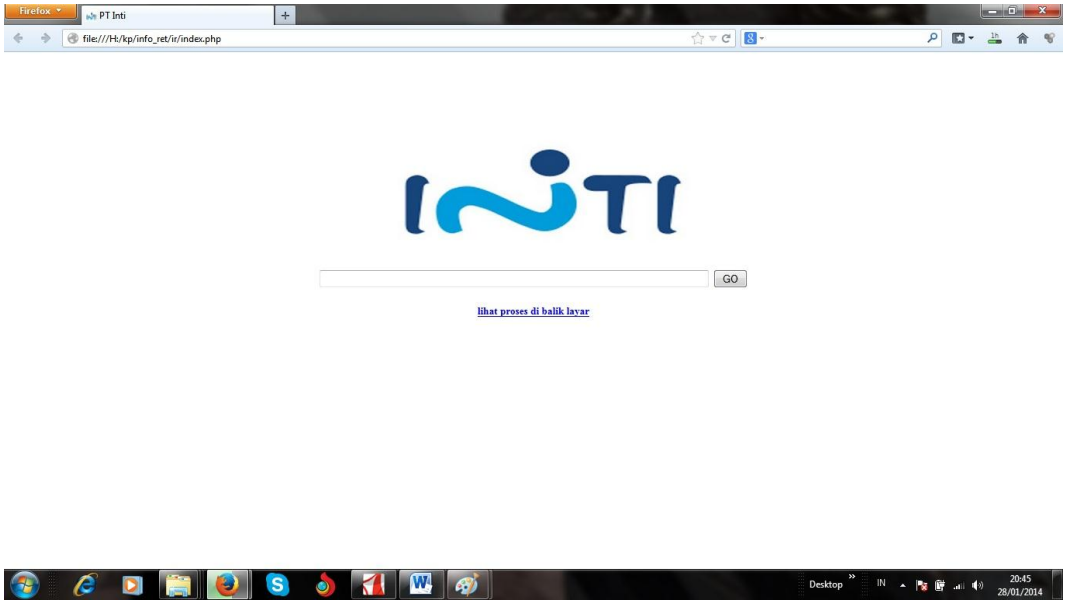
Gambar 3.43 Tampilan Halaman Pembobotan Kata



Gambar 3.44 Tampilan Halaman Retrieval



Gambar 3.45 Halaman Cache



Gambar 3.46 Tampilan Awal Web



Gambar 3.47 Tampilan Saat Melakukan Retrieval