BAB III
ANALISIS DAN PERANCANGAN SISTEM

III.1 Analisis Sistem

Analisis dilakukan dengan tujuan untuk mengetahui dan mengevaluasi seluruh komponen yang berhubungan dengan sistem yang akan dibangun. Tahap analisis merupakan tahap sangat penting karena kesalahan pada tahap ini akan menyebabkan kesalahan pada tahap selanjutnya. Tahap analisis sistem dilakukan dengan cara menguraikan suatu sistem yang utuh kedalam bagian-bagian komponennya dengan tujuan untuk mengetahui dan mengevaluasi permasalahan-permasalahan yang ada sehingga ditemukan kelemahan dan keuntungan pada sistem tersebut.

III.1.1 Analisis Masalah

Metode *Web Content Mining* (WCM) merupakan salah satu metode yang diimplementasikan untuk menemukan informasi dari konten *webpage* suatu situs. Sistem *news aggregator* membutuhkan sebuah pengolahan data informasi dalam menampilkan informasi berita yang berelavan dari situs-situs layanan berita karena sistem *news aggregator* "polos" atau yang hanya mengumpulkan berita saja memiliki kelemahan, yaitu saat menggabungkan semua berita hasil perolehannya dalam satu kumpulan besar dapat mengakibatkan efek *information overload* bagi pengguna karena belum tentu pengguna tertarik dengan semua berita yang terdapat pada *news aggregator*.

Berdasarkan masalah tersebut, diperlukannya metode WCM dengan menggunakan *partitional clustering* K-Means dalam pengolahan sistem *news aggregator* ini agar sistem dapat menampilkan kumpulan informasi berdasarkan inputan kata kunci dari pengguna.

III.1.2 Analisis Data

Dalam sistem ini dibutuhkannya masukan data sebagai bahan untuk proses dalam menjalankan sistem *news aggregator*. Data tersebut adalah data masukkan *query* atau kata kunci dari pengguna serta data dokumen berita yang diambil dari

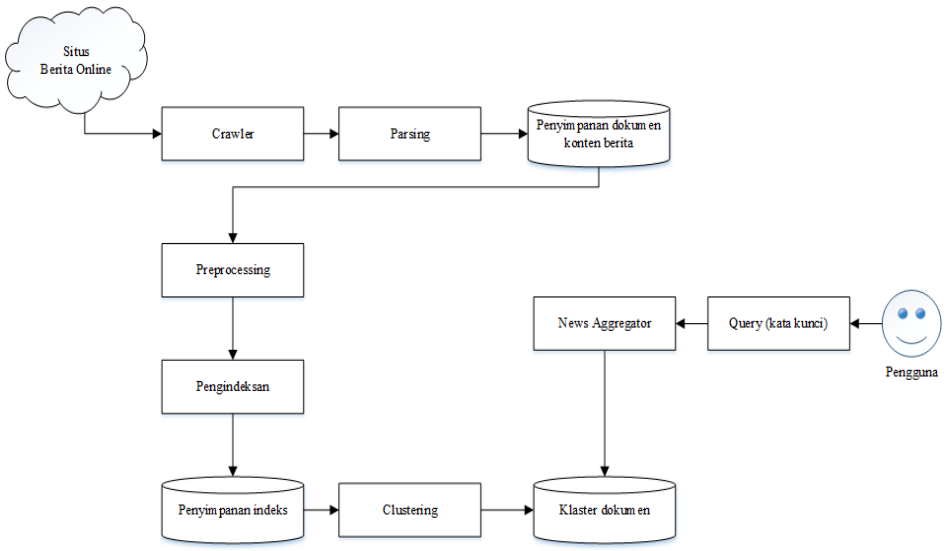
situs-situs berita, dan data-data tersebut berupa data teks. Data-data tersebut akan ke dalam *database*

Tujuan adanya data *query* dengan data dokumen berita dalam sistem ini adalah untuk mengetahui kemiripan atau kesesuaian antara *query* yang dimasukkan oleh pengguna dengan informasi berita yang diambil dari tiga situs berita *online*. Dari kemiripan itu, data dokumen berita dapat dipisahkan ke dalam beberapa kelompok, yaitu kelompok yang paling sesuai atau mirip dengan data *query* dan kelompok yang tidak memiliki kesamaan dengan *query*. Data dokumen berita yang berada di kelompok yang paling sesuai, akan ditampilkan sebagai hasil pencarian ke pengguna. Topik untuk data berita dalam penelitian skripsi ini berupa seputar pendidikan.

III.1.3 Analisis Metode *Web Content Mining*

Metode WCM merupakan sebuah metode untuk mendapatkan infomarsi yang berada di *World Wide Web*. Pada penelitian skripsi ini seperti yang dijelaskan pada sub bab sebelumnya bahwa data yang difokuskan adalah berupa data teks, sehingga metode WCM ini akan digunakan untuk memperoleh informasi berita dari situs-situs berita *online* dan disatukan ke dalam sistem *news aggregator*. Metode WCM sering disebut juga metode *text mining* karena data yang berada di *World Wide Web* sebagian besar berupa teks. Strategi yang digunakan dalam metode WCM ini yang akan diterapkan ke dalam sistem *news aggregator* adalah dengan melakukan pencarian atau dalam penerapan sistemnya berupa *search engine*. Oleh karena itu, di penelitian ini WCM akan menerapkan metode *text mining* dengan teknik *web crawler* untuk mendapatkan informasi beritanya.

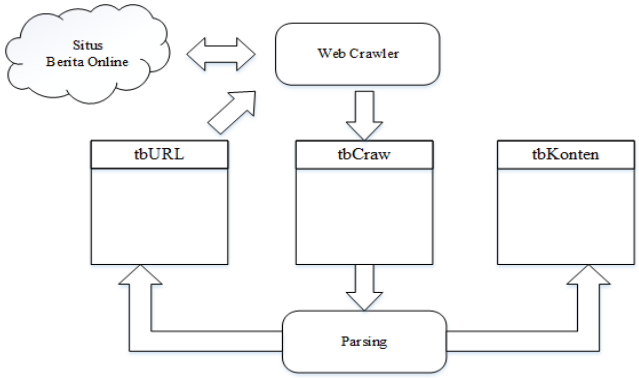
Dalam metode ini memiliki beberapa tahap yang akan dilakukan, mulai dari penelusuran ke situs-situs, tahap *preprocessing* (penghapusan tag HTML, *tokenizing*, *filtering*, *stemming*, *analyzing*) untuk data dokumen berita, dan tahap *clustering* untuk mengelompokkan data berita yang sesuai dengan masukkan *query* dari pengguna.



Gambar III.1 Arsitektur Sistem dengan Metode WCM

III.1.3.1 Tahap *Crawling*

Secara umum, *web crawler* akan diberi URL awal untuk melakukan penelusuran dengan menggunakan metode *Breadth-First Crawler* dan akan melakukan tahap *parsing*. Tahap *parsing* digunakan untuk mengambil *link* yang dari *link* tersebut akan dilakukan proses *crawling* lagi, serta untuk mengambil isi berita yang dari *link* berisi berita tersebut yang selanjutnya dilakukan tahap *preprocessing*.



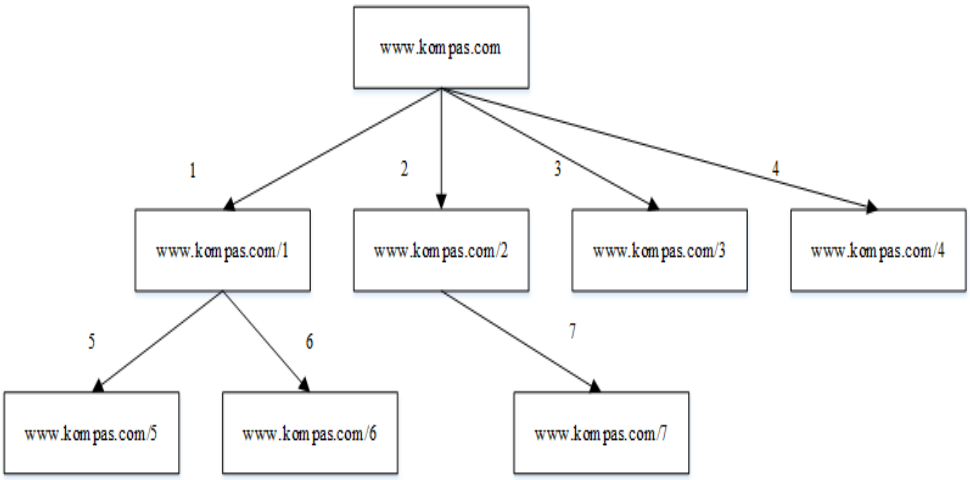
Gambar III.2 Arsitektur Sistem *Web Crawler*

Contoh kasus :

Pada Gambar III.3 merupakan alur dari proses *crawling* yang dimana *link* www.kompas.com sebagai *seed* atau akar. Sebelum melakukan *crawling*, akan ditentukan terlebih dahulu jumlah *depth* atau *max* level yang berguna untuk

membatasi sampai mana proses *crawling* akan terjadi. Berdasarkan Gambar III.3, maka jumlah *depth* yang ditentukan adalah 2 (level 0 sampai level 2). Level 0 (*Top* Level) dimulai dari *seed* atau akar yang telah ditentukan sebelumnya. *Crawler* akan akan melakukan *crawling link* yang ada di level 0 dan *link* yang ditemukan ada 4 *link*. Bila pada level 0 tidak ada *link* yang ditelusuri lagi, maka proses *crawling* selanjutnya dilakukan pada level 1 atau *link – link* yang ditemukan dari level 0.

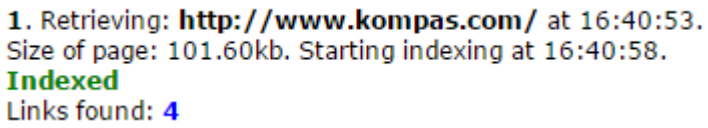
Saat proses *crawling* pada *link* www.kompas.com/1 ditemukan 2 *link*. Selanjutnya dilakukan lagi proses *crawling* yang masih di level 1 pada *link* www.kompas.com/2 dan pada *link* tersebut ditemukan 1 *link*. Proses *crawling* dilanjutkan terus hingga ke *link* www.kompas.com/4 atau *link* terakhir pada level 1, sehingga proses *crawling* selanjutnya dilakukan di level 2. Proses *crawling* sama seperti yang di level 0 dan 1 di mana akan menelusuri *link – link* yang didapat dari level 1. Saat menelusuri *link* dari www.kompas.com/5 sampai www.kompas.com/7 dan ditemukan *link* baru dari *link – link* tersebut, proses *crawling* akan berhenti pada *link* terakhir karena sesuai jumlah *depth* yang ditentukan sebelumnya.



Gambar III.3 Alur *Breadth-First Crawler*

Hasil *crawling* berupa *link – link* tersebut akan di *parsing* terlebih dahulu untuk memisahkan *link* yang berguna untuk proses *crawling* selanjutnya atau

melakukan *crawling* ulang dan untuk mengambil isi berita. Hasil *parsing* tersebut disimpan ke dalam *database*.



Gambar III.4 Hasil Proses *Crawling*

III.1.3.2 Tahap Penghapusan tag HTML

Tahap ini menghapus semua tag *markup* dan format khusus dari dokumen, terutama pada dokumen yang mempunyai banyak tag dan format seperti dokumen XHTML. Untuk penghapusan tag HTML menggunakan fungsi PHP `strip_tags()`. *Pseudocode* : `strip_tags("<b>buku</b>");`. Penghapusan tag HTML ini sebenarnya dilakukan pada saat tahap *parsing*. Berikut contoh penghapusan HTML yang diasumsikan bahwa konten berita telah melalui tahap *crawling* dan *parsing*.

Dokumen berita :

> Prinsipnya, kami siap melaksanakannya penarikan `<b>buku</b>`. Namun, sampai hari ini, tindak lanjut penarikan buku ini belum ada sehingga buku-buku ini masih ada di tangan siswa," kata Leni, Selasa (31/3/2015) siang.

Setelah melakukan penghapusan tag HTML :

> Prinsipnya, kami siap melaksanakannya penarikan buku. Namun, sampai hari ini, tindak lanjut penarikan buku ini belum ada sehingga buku-buku ini masih ada di tangan siswa," kata Leni, Selasa (31/3/2015) siang.

III.1.3.3 Tahap *Tokenizing*

Tujuan dari tahap ini adalah memisahkan atau memecahkan deretan kata dari suatu kalimat, paragraf atau halaman data berita menjadi *token* atau potongan kata tunggal. Tahap ini juga menghilangkan karakter-karakter tertentu seperti

tanda baca dan angka dan mengubah semua *token* ke bentuk huruf kecil (*lowercase*). Karakter tersebut didefinisikan pada tabel berikut :

Tabel III.1 Karakter - karakter yang dihilangkan

No	Karakter
1	%
2	$
3	^
4	&
5	‘
6	“
7	)
8	(
9	\
10	/
11	=
12	.
13	,
14	:
15	-
16	*
17	;
18	!
19	?
20	0
21	1
22	2
23	3
24	4
25	5
26	6
27	7

No	Karakter
28	8
29	9

Contoh kasus :

Diasumsikan ada enam berita atau dokumen dengan sebuah *query* yang sudah melalui tahap penghapusan tag HTML:

D1 = Ungkap Kebocoran Soal UN, Polri Kerja Sama dengan Google Untuk Menghindari Kebocoran Tersebut.

D2 = Surati UGM Karena Soal UN Bocor, Tsaqif Sempat Terima Ancaman.

D3 = Tolak Bocoran Soal UN, Tiga Siswa SMAN 3 Yogyakarta Diberi Penghargaan oleh KPK.

D4 = Buku Berisi Ajaran Garis Keras Masih Dipegang Para Siswa SMA.

D5 = Menteri Anies Kaget Ada Buku Agama Bermuatan Ajaran ISIS.

Maka hasil *tokenizing* dari lima dokumen tersebut :

Tabel III.2 Hasil *Tokenizing*

Dokumen	*Tokenizing*
D1	ungkap – kebocoran – soal – un – polri – kerja – sama – dengan – google – untuk – menghindari – kebocoran - tersebut
D2	surati – ugm – karena – soal – un – bocor – tsaqif – sempat – terima - ancaman
D3	tolak – bocoran - soal – un – tiga – siswa – sman – yogyakarta – diberi – penghargaan – oleh - kpk
D4	buku – berisi – ajaran – garis – keras – masih – dipegang – para – siswa – sma
D5	menteri – anies – kaget – ada – buku – agama – bermuatan – ajaran - isis

Berdasarkan *output* dari proses *tokenizing*, bisa dilihat bahwa ada karakter-karakter yang dihilangkan selama proses *tokenizing*.

III.1.3.4 Tahap *Filtering*

Tahap ini untuk mengambil kata-kata penting atau yang berhubungan dengan kasus masalah dari hasil *tokenizing*. Tahap ini bisa disebut tahap *stopword*

removal karena menghilangkan kata tidak penting pada suatu teks. Jika kata dari proses *tokenizing* ada yang ditemukan di daftar *stopwords*, maka kata tersebut akan dihilangkan atau dianggap sebagai kata tidak penting. Berikut contoh sebagian daftar kata *stopword* :

Tabel III.3 Daftar Kata *Stopword*

No	Kata *Stopword*
1	ada
2	ungkap
3	antara
4	bagi
5	baik
6	sama
7	demikian
8	kembali
9	tersebut
10	secara
11	cukup
12	masih
13	sempat
14	dengan
15	para
16	yaitu
17	hanya
18	harus
19	untuk
20	diberi
21	itu
22	juga
23	tahu
24	karena
25	mulai

No	Kata *Stopword*
26	lagi
27	mengerti
28	pilih
29	oleh
30	yang

Contoh proses *filtering* :

Tabel III.4 Proses *Filtering* / *Stopword Removal*

Dokumen	*Tokenizing*	*Filtering* / *Stopword removal*
D1	ungkap – kebocoran – soal – un – polri – kerja – sama – dengan – google – untuk – menghindari – kebocoran - tersebut	kebocoran – soal – un – polri – kerja – google – menghindari – kebocoran
D2	surati – ugm – karena – soal – un – bocor – tsaqif – sempat – terima - ancaman	surati – ugm – soal – un – bocor – tsaqif – terima - ancaman
D3	tolak – bocoran - soal – un – tiga – siswa – sman – yogyakarta – diberi – penghargaan – oleh - kpk	tolak – bocoran - soal – un – tiga – siswa – sman – yogyakarta – penghargaan - kpk
D4	buku – berisi – ajaran – garis – keras – masih – dipegang – para – siswa – sma	buku – berisi – ajaran – garis – keras – dipegang – siswa – sma
D5	menteri – anies – kaget – ada – buku – agama – bermuatan – ajaran - isis	menteri – anies – kaget – buku – agama – bermuatan – ajaran - isis

III.1.3.5 Tahap *Stemming*

Tahap *stemming* merupakan proses 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 telah melalui tahap *tokenizing* dengan daftar kamus *stem*.

Contoh *stemming* yang telah melalui tahap *filtering* :

Tabel III.5 Proses *Stemming*

Dokumen	*Filtering* / *Stopword removal*	*Stemming*

Dokumen	*Filtering / Stopword removal*	*Stemming*
D1	kebocoran – soal – un – polri – kerja – google – menghindari – kebocoran	bocor – soal – un – polri – kerja – google – hindar – bocor
D2	surati – ugm – soal – un – bocor – tsaqif – terima - ancaman	surati – ugm – soal – un – bocor – tsaqif – terima - ancam
D3	tolak – bocoran - soal – un – tiga – siswa – sman – yogyakarta – penghargaan - kpk	tolak – bocor - soal – un – tiga – siswa – sman – yogyakarta – harga - kpk
D4	buku – berisi – ajaran – garis – keras – dipegang – siswa – sma	buku – isi – ajar – garis – keras – pegang – siswa – sma
D5	menteri – anies – kaget – buku – agama – bermuatan – ajaran - isis	menteri – anies – kaget – buku – agama – muat – ajar - isis

III.1.3.6 Tahap *Analyzing*

Tahap *analyzing* atau *weighting* (pembobotan) merupakan tahap untuk menentukan tingkat similaritas antar kalimat. Tahap ini menggunakan rumus TF-IDF. Dari hasil perhitungan TF-IDF ini akan dibentuk suatu vektor antara dokumen dengan kata yang kemudian untuk kesamaan antar dokumen dengan klaster akan ditentukan oleh sebuah klaster *centroid*. Berikut rumus TF-IDF yang dinormalisasikan sebagai berikut [4] :

$$W_{td} = \frac{tf_{td} * \left(\log\left({}^{N}/_{df_t}\right)+1\right)}{\sqrt{\sum_{k=1}^{t}(tf_{td})^2 \times \left[\left(\log\left({}^{N}/_{df_t}\right)+1\right)\right]^2}}$$ Persamaan III-1

Keterangan :

Wtd : Bobot kata/token terhadap dokumen

tftd : Jumlah kemunculan *term* di dokumen

N : Jumlah dokumen yang ada di *database*

dft : Jumlah dokumen yang mengandung kata/token

Tabel III.6 Data TF dan DF

Kata/ Token	TF					DF
	D1	D2	D3	D4	D5	
agama	0.00	0.00	0.00	0.00	1.00	1

Kata/ Token	TF					DF
	D1	D2	D3	D4	D5	
ajar	0.00	0.00	0.00	1.00	1.00	2
ancam	0.00	1.00	0.00	0.00	0.00	1
anies	0.00	0.00	0.00	0.00	1.00	1
bocor	2.00	1.00	1.00	0.00	0.00	3
buku	0.00	0.00	0.00	1.00	1.00	2
garis	0.00	0.00	0.00	1.00	0.00	1
google	1.00	0.00	0.00	0.00	0.00	1
harga	0.00	0.00	1.00	0.00	0.00	1
hindar	1.00	0.00	0.00	0.00	0.00	1
isi	0.00	0.00	0.00	1.00	0.00	1
isis	0.00	0.00	0.00	0.00	1.00	1
kaget	0.00	0.00	0.00	0.00	1.00	1
keras	0.00	0.00	0.00	1.00	0.00	1
kerja	1.00	0.00	0.00	0.00	0.00	1
kpk	0.00	0.00	1.00	0.00	0.00	1
menteri	0.00	0.00	0.00	0.00	1.00	1
muat	0.00	0.00	0.00	0.00	1.00	1
pegang	0.00	0.00	0.00	1.00	0.00	1
polri	1.00	0.00	0.00	0.00	0.00	1
siswa	0.00	0.00	1.00	1.00	0.00	2
sma	0.00	0.00	0.00	1.00	0.00	1
sman	0.00	0.00	1.00	0.00	0.00	1
soal	1.00	1.00	1.00	0.00	0.00	3
surati	0.00	1.00	0.00	0.00	0.00	1
terima	0.00	1.00	0.00	0.00	0.00	1
tiga	0.00	0.00	1.00	0.00	0.00	1
tolak	0.00	0.00	1.00	0.00	0.00	1
tsaqif	0.00	1.00	0.00	0.00	0.00	1
ugm	0.00	1.00	0.00	0.00	0.00	1

Kata/ Token	TF					DF
	D1	D2	D3	D4	D5	
un	1.00	1.00	1.00	0.00	0.00	3
yogyakarta	0.00	0.00	1.00	0.00	0.00	1

Contoh menghitung TF-IDF token “bocor” pada D1 dengan rumus TF-IDF : N (jumlah dokumen) = 5, tf1 = 2, dan df1 = 4, sehingga dapat dihitung :

$$idf_1 = \left(\log {}^{5}/_{4}\right) + 1 = 1.0969$$

$$W_{11} = \frac{2 \times 1.0969}{\sqrt{((2^2 \times [1.0969]^2) + (1^2 \times [1.0969]^2) + \cdots + (1^2 \times [1.0969]^2))}}$$

$$W_{11} = \frac{2.1938}{3.4687}$$

$$W_{11} = 0.6325$$

Sehingga pembobotan untuk semua kata di setiap dokumennya :

Tabel III.7 Hasil TF-IDF

Kata/ Token	TF					DF	TF-IDF				
	D1	D2	D3	D4	D5		WD1	WD2	WD3	WD4	WD5
agama	0.00	0.00	0.00	0.00	1.00	1	0.00	0.00	0.00	0.00	0.3536
ajar	0.00	0.00	0.00	1.00	1.00	2	0.00	0.00	0.00	0.3536	0.3536
ancam	0.00	1.00	0.00	0.00	0.00	1	0.00	0.3536	0.00	0.00	0.00
anies	0.00	0.00	0.00	0.00	1.00	1	0.00	0.00	0.00	0.00	0.3536
bocor	2.00	1.00	1.00	0.00	0.00	3	0.6325	0.3536	0.3162	0.00	0.00
buku	0.00	0.00	0.00	1.00	1.00	2	0.00	0.00	0.00	0.3536	0.3536
garis	0.00	0.00	0.00	1.00	0.00	1	0.00	0.00	0.00	0.3536	0.00
google	1.00	0.00	0.00	0.00	0.00	1	0.3162	0.00	0.00	0.00	0.00
harga	0.00	0.00	1.00	0.00	0.00	1	0.00	0.00	0.3162	0.00	0.00
hindar	1.00	0.00	0.00	0.00	0.00	1	0.3162	0.00	0.00	0.00	0.00
isi	0.00	0.00	0.00	1.00	0.00	1	0.00	0.00	0.00	0.3536	0.00

Kata/ Token	TF					DF	TF-IDF				
	D1	D2	D3	D4	D5		WD1	WD2	WD3	WD4	WD5
isis	0.00	0.00	0.00	0.00	1.00	1	0.00	0.00	0.00	0.00	0.3536
kaget	0.00	0.00	0.00	0.00	1.00	1	0.00	0.00	0.00	0.00	0.3536
keras	0.00	0.00	0.00	1.00	0.00	1	0.00	0.00	0.00	0.3536	0.00
kerja	1.00	0.00	0.00	0.00	0.00	1	0.3162	0.00	0.00	0.00	0.00
kpk	0.00	0.00	1.00	0.00	0.00	1	0.00	0.00	0.3162	0.00	0.00
menteri	0.00	0.00	0.00	0.00	1.00	1	0.00	0.00	0.00	0.00	0.3536
muat	0.00	0.00	0.00	0.00	1.00	1	0.00	0.00	0.00	0.00	0.3536
pegang	0.00	0.00	0.00	1.00	0.00	1	0.00	0.00	0.00	0.3536	0.00
polri	1.00	0.00	0.00	0.00	0.00	1	0.3162	0.00	0.00	0.00	0.00
siswa	0.00	0.00	1.00	1.00	0.00	2	0.00	0.00	0.3162	0.3536	0.00
sma	0.00	0.00	0.00	1.00	0.00	1	0.00	0.00	0.00	0.3536	0.00
sman	0.00	0.00	1.00	0.00	0.00	1	0.00	0.00	0.3162	0.00	0.00
soal	1.00	1.00	1.00	0.00	0.00	3	0.3162	0.3536	0.3162	0.00	0.00
surati	0.00	1.00	0.00	0.00	0.00	1	0.00	0.3536	0.00	0.00	0.00
terima	0.00	1.00	0.00	0.00	0.00	1	0.00	0.3536	0.00	0.00	0.00
tiga	0.00	0.00	1.00	0.00	0.00	1	0.00	0.00	0.3162	0.00	0.00
tolak	0.00	0.00	1.00	0.00	0.00	1	0.00	0.00	0.3162	0.00	0.00
tsaqif	0.00	1.00	0.00	0.00	0.00	1	0.00	0.3536	0.00	0.00	0.00
ugm	0.00	1.00	0.00	0.00	0.00	1	0.00	0.3536	0.00	0.00	0.00
un	1.00	1.00	1.00	0.00	0.00	3	0.3162	0.3536	0.3162	0.00	0.00
yogyakarta	0.00	0.00	1.00	0.00	0.00	1	0.00	0.00	0.3162	0.00	0.00

III.1.3.7 Tahap *Clustering*

Tahap *clustering* ini menggunakan Algoritma K-Means. Pada tahap ini akan dicari *centroid* yang terdekat berdasarkan nilai *euclidean distance* yang dicari dari setiap dokumen. Adapun contoh beserta langkah-langkah tahap *clustering* dengan menerapkan algoritma K-Means dengan sebagai berikut :

1. Menentukan jumlah *cluster* dan *centroid* awal. Banyaknya *cluster* harus lebih kecil dari banyaknya data (k < n). Berdasarkan pada contoh kasus

sebelumnya nilai k yang digunakan adalah 3 dan *centroid* awal yang terpilih secara random (D1, D4 dan D5). Maka *centroid* pertama pada *cluster* C1 diinisialisasikan pada D1, *centroid* kedua pada *cluster* C2 berada pada D4 dan *centroid* pada *cluster* C3 berada pada D5.

2. Langkah selanjutnya adalah mencari *centroid* terdekat dari setiap dokumen. Pada penelitian ini akan menerapkan pendekatan berdasarkan *euclidean distance* yang digunakan untuk mencari jarak antara masing-masing dokumen dengan *centroid*. Adapun rumus pengukuran *euclidean distance* yang didefinisikan sebagai berikut :

$$d_{(i,j)} = \sqrt{\left(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \cdots + |x_{in} - x_{jn}|^2\right)} \quad \text{Persamaan III-2}$$

Keterangan :

$d_{(i,j)}$: jarak dokumen ke i ke dokumen ke j

x_{in} : kata ke n di dokumen ke-i

x_{jn} : kata ke n di dokumen ke-j

Contoh perhitungan mencari nilai *euclidean distance* antara *centroid* C1 dengan dokumen D2 berdasarkan pada Tabel III.7 :

$$d_{(D2,centroid1)} = \sqrt{(|0-0|^2 + |0-0|^2 + |0.3536-0|^2 + \cdots + |0-0|^2)}$$

$$d_{(D2,centroid1)} = \sqrt{1.1056}$$

$$d_{(D2,centroid1)} = 1.0515$$

Maka dapat ditentukan *euclidean distance* dari masing-masing dokumen dengan *centroid* yang telah ditentukan dengan hasilnya diperlihatkan pada Tabel III.8 berikut.

Tabel III.8 Hasil *Euclidean Distance*

Dokumen	*Centroid*	Jarak (d)
D2	C1	**1.0515**
D2	C2	1.4142
D2	C3	1.4142

Dokumen	*Centroid*	Jarak (d)
D3	C1	**1.0954**
D3	C2	1.3328
D3	C3	1.4142

Pada tabel di atas, dapat diketahui bahwa dokumen D2 dan D3 masuk sebagai *cluster* C1 karena memiliki jarak paling kecil daripada jarak pada *centroid* lain.

3. Langkah selanjutnya menghitung ulang untuk menentukan *centroid* baru dari setiap *cluster*. Adapun rumus untuk menentukan *centroid* baru :

$$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} X_{ik}$$ Persamaan III-3

Keterangan :

M_k : Nilai *centroid* dari suatu *cluster*

n_k : Jumlah dokumen yang berada dalam satu *cluster*

X_{ik} : Nilai vektor dari sampel dokumen ke-i yang termasuk ke dalam *cluster* Ck

Dari langkah sebelumnya, telah diketahui bahwa dokumen D2 dan D3 termasuk ke dalam *cluster* C1 sehingga dalam *cluster* C1 sekarang memiliki tiga dokumen, yaitu D1, D2 dan D3, sedangkan dalam *cluster* C2 terdapat dokumen D4 dan di *cluster* C3 terdapat dokumen D5. Berdasarkan pada *cluster* C1 dapat ditentukan *centroid* baru dengan menggunakan rumus di atas.

Tabel III.9 Hasil perhitungan *centroid* baru pada C1

Kata/ Token	D1	D2	D3	$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} X_{ik}$
agama	0.00	0.00	0.00	0.0000
ajar	0.00	0.00	0.00	0.0000

Kata/ Token	D1	D2	D3	$M_k = \left(\frac{1}{n_k}\right)\sum_{i=1}^{n_k} X_{ik}$
ancam	0.00	0.3536	0.00	0.1179
anies	0.00	0.00	0.00	0.0000
bocor	0.6325	0.3536	0.3162	0.4341
buku	0.00	0.00	0.00	0.0000
garis	0.00	0.00	0.00	0.0000
google	0.3162	0.00	0.00	0.1054
harga	0.00	0.00	0.3162	0.1054
hindar	0.3162	0.00	0.00	0.1054
isi	0.00	0.00	0.00	0.0000
isis	0.00	0.00	0.00	0.0000
kaget	0.00	0.00	0.00	0.0000
keras	0.00	0.00	0.00	0.0000
kerja	0.3162	0.00	0.00	0.1054
kpk	0.00	0.00	0.3162	0.1054
menteri	0.00	0.00	0.00	0.0000
muat	0.00	0.00	0.00	0.0000
pegang	0.00	0.00	0.00	0.0000
polri	0.3162	0.00	0.00	0.1054
siswa	0.00	0.00	0.3162	0.1054
sma	0.00	0.00	0.00	0.0000
sman	0.00	0.00	0.3162	0.1054
soal	0.3162	0.3536	0.3162	0.3287
surati	0.00	0.3536	0.00	0.1179

Kata/ Token	D1	D2	D3	$M_k = \left(\frac{1}{n_k}\right)\sum_{i=1}^{n_k} X_{ik}$
terima	0.00	0.3536	0.00	0.1179
tiga	0.00	0.00	0.3162	0.1054
tolak	0.00	0.00	0.3162	0.1054
tsaqif	0.00	0.3536	0.00	0.1179
ugm	0.00	0.3536	0.00	0.1179
un	0.3162	0.3536	0.3162	0.3287
yogyakarta	0.00	0.00	0.3162	0.1054

4. Langkah terakhir adalah menghitung apakah *centroid* lama sama dengan *centroid* baru. Untuk itu akan digunakan kembali rumus *euclidean distance* antara *centroid* baru dengan *centroid* lama. Dengan melakukan perhitungan *euclidean* maka dihasilkan :

Tabel III.10 Hasil Perhitungan *centroid* lama dengan *centroid* baru

Centroid	
C1 lama	C1 Baru
0.00	0.0000
0.00	0.0000
0.00	0.1179
0.00	0.0000
0.6325	0.4341
0.00	0.0000
0.00	0.0000
0.3162	0.1054
0.00	0.1054

Centroid	
0.3162	0.1054
0.00	0.0000
0.00	0.0000
0.00	0.0000
0.00	0.0000
0.3162	0.1054
0.00	0.1054
0.00	0.0000
0.00	0.0000
0.00	0.0000
0.3162	0.1054
0.00	0.1054
0.00	0.0000
0.00	0.1054
0.3162	0.3287
0.00	0.1179
0.00	0.1179
0.00	0.1054
0.00	0.1054
0.00	0.1179
0.00	0.1179
0.3162	0.3287
0.00	0.1054
Jarak	0.6039

Dari hasil tersebut maka akan dilakukan iterasi lagi (kembali ke langkah 2) hingga semua dokumen masuk ke *cluster* yang ditentukan. Dari contoh kasus sebelumnya, maka didapatkan hasil *cluster* terhadap lima dokumen :

Tabel III.11 *Cluster* dokumen

Cluster C1	*Cluster* C2	*Cluster* C3
D1	D4	D5
D2		
D3		

III.1.3.8 Keluaran Sistem *News Aggregator*

Keluaran yang diharapkan dari sistem *news aggregator* dengan menerapkan metode WCM ini adalah berupa hasil pencarian berita yang telah dikelompokkan berdasarkan topik yang sama dari inputan kata kunci pengguna.

III.2 Spesifikasi Perangkat Lunak

Spesifikasi perangkat lunak yang dibangun dibagi menjadi dua kebutuhan, yaitu kebutuhan nonfungsional dan fungsional. Spesifikasi kebutuhan peraangkat lunak dapat dilihat pada Tabel III.12 untuk SKPL fungsional dan Tabel III.13 untuk SKPL nonfungsional.

Tabel III.12 Spesifikasi Kebutuhan Fungsional

Nomor	Spesifikasi Kebutuhan Fungsional
SKPL-F-001	Sistem dapat menambah situs berita
SKPL-F-002	Sistem dapat mengedit situs berita
SKPL-F-003	Sistem dapat mengahapus situs berita
SKPL-F-004	Sistem dapat *crawling* situs berita
SKPL-F-005	Sistem dapat *crawling* semua situs berita
SKPL-F-006	Sistem dapat menampilkan hasil proses *crawling*
SKPL-F-007	Sistem dapat menambah kategori
SKPL-F-008	Sistem dapat mengedit kategori
SKPL-F-009	Sistem dapat menghapus kategori
SKPL-F-010	Sistem dapat menampilkan semua informasi berita dari hasil *crawling* situs berita
SKPL-F-011	Sistem dapat menampilkan informasi berita dari hasil *crawling* situs berita berdasarkan kategori situs berita
SKPL-F-012	Sistem dapat mencari dan menampilkan informasi judul berita dari dari hasil *crawling* situs berita berdasarkan kata kunci/*query*

Tabel III.13 Spesifikasi Kebutuhan NonFungsional

Nomor	Spesifikasi Kebutuhan NonFungsional
SKPL-NF-001	Sistem yang dibangun menggunakan *database* MySQL versi 5.6.
SKPL-NF-002	Sistem dapat berjalan dengan baik jika ada koneksi internet.
SKPL-NF-003	Tidak ada pengguna khusus untuk menggunakan sistem ini.
SKPL-NF-004	Sistem ini membutuhkan *Microsoft Windows* sebagai sistem operasi dan XAMPP untuk *web server*.

III.3 Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional menggambarkan kebutuhan luar sistem yang diperlukan untuk menjalakan *prototype* yang dibangun. Adapun analisis kebutuhan non fungsional untuk menjalankan *prototype news aggregator* ini meliputi, analisis kebutuhan perangkat lunak, analisis kebutuhan perangkat keras, dan analisis kebutuhan pengguna yang untuk menjalankan sistem *prototype* ini. Analisis kebutuhan non fungsional bertujuan agar *prototype* yang dibangun dapat digunakan sesuai dengan kebutuhan.

III.3.1.1 Analisis Kebutuhan Perangkat Lunak

Perangkat lunak yang dibutuhkan dalam membangun dan menjalankan sistem *prototype* ini antara lain :

Tabel III.14 Spesfikasi Perangkat Lunak

No	Perangkat Lunak	Spesifikasi
1	Sistem Operasi	*Microsoft Windows*
2	*Database & Web Server*	*MySQL* 5.6 & *Xamp Server*
3	Editor PHP	*Adobe Dreamweaver*
4	*Web Browser*	Opera

III.3.1.2 Analisis Kebutuhan Perangkat Keras

Untuk menjalankan sistem *news aggregator* ini maka dibutuhkan perangkat keras yang dapat mendukung proses kerja dari sistem *news aggregator*. Adapun spesifikasi agar sistem ini dapat berjalan dengan semestinya bisa dilihat pada Tabel III.15 :

Tabel III.15 Spesfikasi Perangkat Keras

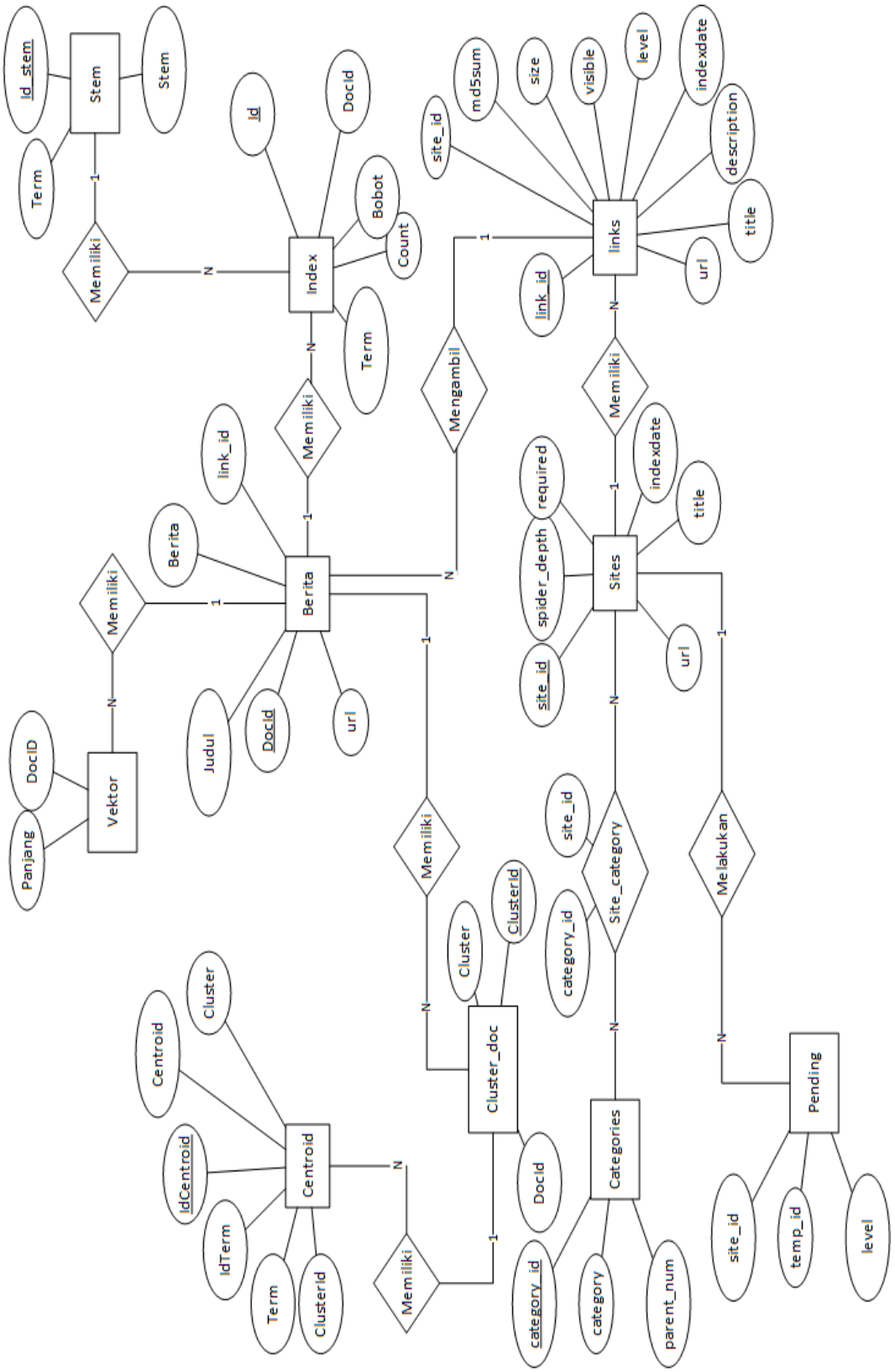
No	Perangkat Keras	Spesifikasi
1	Prosesor	Kecepatan 1.50 GHz
2	Monitor	Monitor 14”, resolusi 1366x768
3	VGA	On-Board 792 MB
4	Keyboard	Standard
5	Mouse	Standard
6	Memori	2 GB

III.3.1.3 Analisis Kebutuhan Pengguna

Pada sistem *news aggregator* ini tidak ada pengguna khusus untuk menjalankan sistem tersebut. Pengguna di sini hanya mengatur sistem *news aggregator* untuk *crawling* dan memanfaatkan fungsional pencarian untuk mencari berita.

III.4 Analisis Basis Data

Pemodelan analisis basis data dalam membangun sistem ini menggunakan *Entity Relationship Diagram* (ERD). Berikut adalah ERD dan kardinalitasnya yang ditunjukkan pada Gambar III.5.



Gambar III.5 ERD Sistem yang Dibangun

III.5 Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional ini merupakan analisis tentang hal – hal yang dilakukan sistem ketika diimplemntasikan. Analisis kebutuhan ini untuk menentukan masukan yang dibutuhkan sistem, proses yang digunakan untuk mengolah masukan menjadi keluaran, dan keluaran yang akan dihasilkan oleh sistem. Analisis sistem yang dilakukan menggunakan *tools* UML yang memiliki tahapan analisis sistem yang terdiri dari *use case diagram*, *use case scenario*, *activity diagram*, *class diagram*, dan *sequence diagram*.

III.5.1.1 *Use Case Diagram*

Sasaran pemodelan *use case* diantaranya adalah mendefinisikan gambaran umum dari sebuah sistem sehingga pengguna mengetahu mengenai sistem yang akan dibangun. *Use case diagram* untuk proses keseluruhan sistem *news aggregator* ini dapat dilihat pada Gambar III.6.



Gambar III.6 *Use Case Diagram News Aggregator*

Definisi *use case* dari pembangunan *prototype* ini dapat dilihat pada Tabel III.16.

Tabel III.16 Definisi *Use Case*

No	*Use Case*	Deskripsi
1	Penyajian Tambah Situs Berita	Proses untuk menambah situs berita.
2	Penyajian Edit Situs Berita	Proses untuk mengedit situs berita.
3	Penyajian Hapus Situs Berita	Proses untuk menghapus situs berita.
4	Penyajian *Crawling* Situs Berita	Proses untuk *crawling* situs berita.
5	Penyajian *Crawling* Semua Situs Berita	Proses untuk *crawling* semua situs berita.
6	Penyajian Hasil Proses *Crawling*	Proses untuk menampilkan hasil proses *crawling*.

No	*Use Case*	Deskripsi
7	Penyajian Tambah Kategori	Proses untuk menambah kategori.
8	Penyajian Hapus Kategori	Proses untuk menghapus kategori.
9	Penyajian Edit Kategori	Proses untuk mengedit kategori.
10	Penyajian Informasi Semua Berita	Proses untuk menampilkan informasi semua berita dari hasil *crawling*.
11	Penyajian Pencarian Informasi Berita	Proses untuk mencari dan melihat hasil pencarian informasi berita.berdasarkan kata kunci/*query*
12	Penyajian Informasi Berita Berdasarkan Kategori	Proses untuk menampilkan informasi berita dari hasil *crawling* berdasarkan kategori.

III.5.1.2 *Use Case Scenario*

Use case scenario mendeskripsikan urutan langkah – langkah dalam proses bisnis bisnis baik yang dilakukan aktor terhadap sistem maupun yang dilakukan oleh sistem terhadap aktor. Berdasarkan *use case diagram* pada Gambar III.6 maka *use case scenario* untuk sistem yang dibangun akan dijelaskan sebagai berikut :

Tabel III.17 *Use Case Scenario* Penyajian Tambah Situs Berita

Use Case Name	Penyajian Tambah Situs Berita	
Goal In Context	Pengguna dapat menambah situs berita ke sistem	
Precondition	-	
Description	Proses ini digunakan untuk menambah situs berita	
Related Use Case	Penyajian Tambah Kategori	
Successful End Condition	Sistem berhasil menambah situs berita	
Failed End Condition	Sistem tidak berhasil menambah situs berita	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses tambah situs berita	
Main Flow	**Step**	**Action**
	1.	Pilih Tambah Situs
	2.	Memasukkan URL dan judul
	3.	Pilih jenis kategori (*optional*)
	4.	Klik tombol Tambah
	5.	Sistem menyimpan data tambah situs ke *database*
Extension	3.1	Jika kategori belum ada, maka pilihan jenis kategori tidak muncul

Tabel III.18 *Use Case Scenario* Penyajian Edit Situs Berita

Use Case Name	Penyajian Edit Situs Berita
Goal In Context	Pengguna dapat mengedit situs berita
Precondition	Situs berita telah ditambah
Description	Proses ini digunakan untuk mengedit situs berita

Related Use Case	Penyajian Tambah Situs Berita	
Successful End Condition	Sistem berhasil mengedit situs berita	
Failed End Condition	Sistem tidak berhasil mengedit situs berita	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses edit situs berita	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Pilihan
	2.	Pilih Edit
	3.	Pengguna memasukkan data yang diedit
	4.	Klik tombol Update
	5.	Sistem meng-*update* data situs ke *database*

Tabel III.19 *Use Case Scenario* Penyajian Hapus Situs Berita

Use Case Name	Penyajian Hapus Situs Berita	
Goal In Context	Pengguna dapat menghapus situs berita	
Precondition	Situs berita telah ditambah	
Description	Proses ini digunakan untuk menghapus situs berita	
Related Use Case	Penyajian Tambah Situs Berita	
Successful End Condition	Sistem berhasil menghapus situs berita	
Failed End Condition	Sistem tidak berhasil menghapus situs berita	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses hapus situs berita	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Pilihan
	2.	Pilih Hapus
	3.	Sistem menampilkan pesan validasi hapus situs berita
	4.	Sistem menghapus situs berita dari *database*
Extension	3.1.	Jika memilih Cancel, situs berita tidak dihapus

Tabel III.20 *Use Case Scenario* Penyajian *Crawling* Situs Berita

Use Case Name	Penyajian *Crawling* Situs Berita	
Goal In Context	Sistem dapat *crawling* situs berita	
Precondition	Situs berita telah ditambah	
Description	Proses ini digunakan untuk *crawling* situs berita	
Related Use Case	Penyajian Tambah Situs Berita	
Successful End Condition	Sistem berhasil *crawling* situs berita	
Failed End Condition	Sistem tidak berhasil *crawling* situs berita	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses *crawling* situs berita	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Pilihan
	2.	Pilih *Crawling*
	3.	Memasukkan nilai *depth*

	4.	*Checklist* / *Unchecklist* Reindeks (*optional*)
	5.	Klik tombol Mulai Crawling
	6.	Sistem menampilkan halaman proses *crawling*

Use Case Name	Penyajian *Crawling* Semua Situs Berita	
Goal In Context	Sistem dapat *crawling* semua situs berita	
Precondition	Situs berita telah ditambah minimal 1 situs atau lebih	
Description	Proses ini digunakan untuk *crawling* semua situs berita	
Related Use Case	Penyajian Tambah Situs Berita	
Successful End Condition	Sistem berhasil *crawling* semua situs berita	
Failed End Condition	Sistem tidak berhasil *crawling* semua situs berita	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses *crawling* semua situs berita	
Main Flow	**Step**	**Action**
	1.	Pilih *Crawling* Semua
	2.	Sistem menampilkan halaman proses *crawling*

Tabel III.21 *Use Case Scenario* Penyajian *Crawling* Semua Situs Berita

Tabel III.22 *Use Case Scenario* Penyajian Hasil Proses *Crawling*

Use Case Name	Penyajian Hasil Proses *Crawling*	
Goal In Context	Sistem dapat menampilkan hasil proses *crawling*	
Precondition	Situs berita telah ditambah dan telah memasukkan jumlah *depth*	
Description	Proses ini digunakan untuk menampilkan hasil proses *crawling*	
Related Use Case	Penyajian Tambah Situs Berita, Penyajian *Crawling* Situs Berita, Penyajian *Crawling* Semua Situs Berita	
Successful End Condition	Sistem berhasil menampilkan proses hasil *crawling*	
Failed End Condition	Sistem tidak berhasil menampilkan hasil proses *crawling*	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses *crawling* situs atau semua situs berita	
Main Flow	**Step**	**Action**
	1.	Sistem menampilkan halaman proses *crawling*
	2.	Sistem melakukan indeks situs dan URL
	3.	Sistem melakukan indeks berita
	4.	Sistem melakukan indeks *preprocessing* dan pembobotan
	5.	Sistem melakukan indeks kluster

Tabel III.23 *Use Case Scenario* Penyajian Tambah Kategori

Use Case Name	Penyajian Tambah Kategori
Goal In Context	Pengguna dapat menambah kategori ke sistem

Precondition	-	
Description	Proses ini digunakan untuk menambah kategori	
Related Use Case	-	
Successful End Condition	Sistem berhasil menambah kategori	
Failed End Condition	Sistem tidak berhasil menambah kategori	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses tambah kategori	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Kategori
	2.	Klik Tambah Kategori
	3.	Memasukkan nama kategori
	4.	Klik tombol Tambah
	5.	Sistem menyimpan data tambah kategori ke *database*

Tabel III.24 *Use Case Scenario* Penyajian Edit Kategori

Use Case Name	Penyajian Edit Kategori	
Goal In Context	Pengguna dapat mengedit kategori	
Precondition	Kategori telah ditambah	
Description	Proses ini digunakan untuk mengedit kategori	
Related Use Case	Penyajian Tambah Kategori	
Successful End Condition	Sistem berhasil mengedit kategori	
Failed End Condition	Sistem tidak berhasil mengedit kategori	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses edit kategori	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Kategori
	2.	Pilih Edit
	3.	Pengguna memasukkan data yang diedit
	4.	Klik tombol Update
	5.	Sistem menyimpan data edit kategori ke *database*

Tabel III.25 *Use Case Scenario* Penyajian Hapus Kategori

Use Case Name	Penyajian Hapus Kategori	
Goal In Context	Pengguna dapat menghapus kategori	
Precondition	Kategori telah ditambah	
Description	Proses ini digunakan untuk menghapus kategori	
Related Use Case	Penyajian Tambah Kategori	
Successful End Condition	Sistem berhasil menghapus kategori	
Failed End Condition	Sistem tidak berhasil menghapus kategori	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses hapus kategori	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Kategori
	2.	Pilih Hapus

	3.	Sistem menampilkan pesan validasi hapus kategori
	4.	Sistem menghapus situs berita dari *database*
Extension	3.1.	Jika memilih Cancel, kategori tidak dihapus

Tabel III.26 *Use Case Scenario* Penyajian Informasi Semua Berita

Use Case Name	Penyajian Informasi Semua Berita	
Goal In Context	Sistem dapat menampilkan semua informasi berita	
Precondition	Situs berita telah di *crawling* dan tersimpan dalam *database*	
Description	Proses ini digunakan untuk menampilkan informasi semua berita dari hasil *crawling*	
Related Use Case	Penyajian Hasil Proses *Crawling*	
Successful End Condition	Sistem berhasil menampilkan semua informasi berita	
Failed End Condition	Sistem tidak berhasil menampilkan semua informasi berita	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses menampilkan infromasi semua berita	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Aggregator
	2.	Sistem menampilkan halaman informasi semua berita

Tabel III.27 *Use Case Scenario* Penyajian Informasi Berita Berdasarkan Kategori

Use Case Name	Penyajian Informasi Berita Berdasarkan Kategori	
Goal In Context	Sistem dapat menampilkan informasi berita berdasarkan kategori	
Precondition	Situs berita telah di *crawling* dan tersimpan dalam *database*	
Description	Proses ini digunakan untuk menampilkan informasi semua berita dari hasil *crawling* berdasarkan kategori	
Related Use Case	Penyajian Hasil Proses *Crawling*	
Successful End Condition	Sistem berhasil menampilkan informasi berita berdasarkan kategori	
Failed End Condition	Sistem tidak berhasil menampilkan informasi berita berdasarkan kategori	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses menampilkan infromasi berita berdasarkan kategori	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Aggregator
	2.	Sistem menampilkan halaman informasi semua berita
	3.	Pilih salah satu kategori pada *sidebar*
	4.	Sistem menampilkan informasi berita berdasarkan kategori

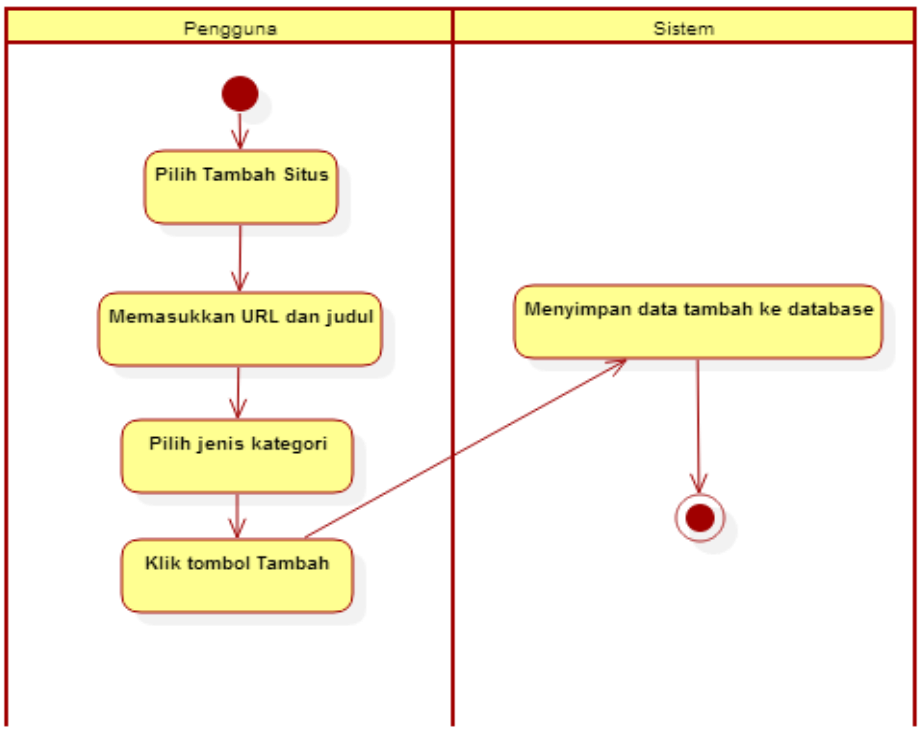
Tabel III.28 Penyajian Pencarian Informasi Berita

Use Case Name	Penyajian Pencarian Informasi Berita

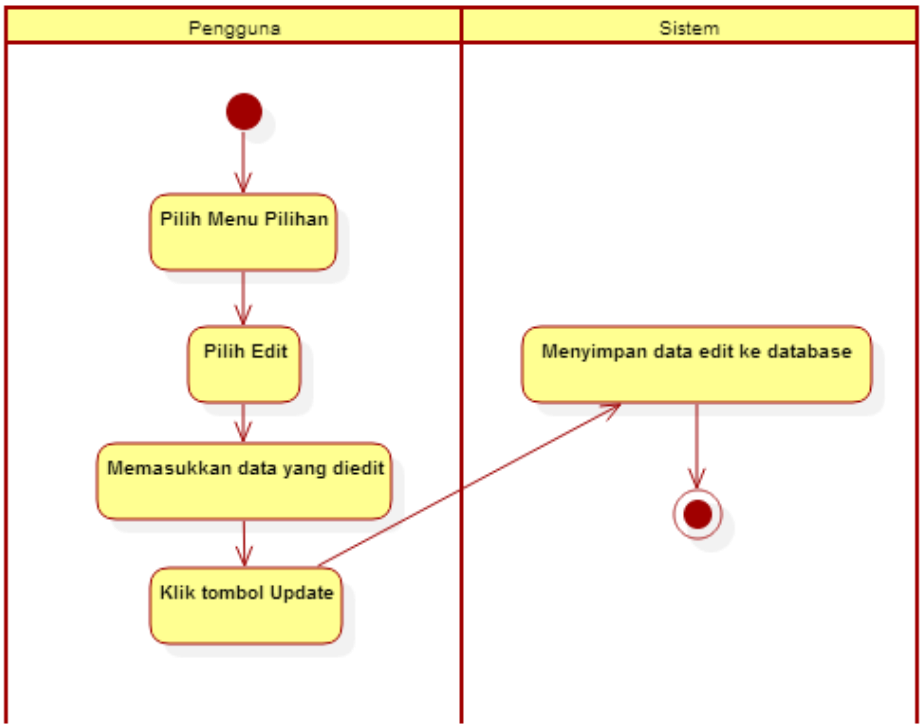
Goal In Context	Pengguna dapat mencari dan melihat informasi berita dengan memasukkan kata kunci/*query*	
Precondition	Situs berita telah di *crawling* dan tersimpan dalam *database*	
Description	Proses ini digunakan untuk mencari dan melihat hasil pencarian informasi berita.berdasarkan kata kunci/*query*	
Related Use Case	Penyajian Hasil Proses *Crawling*	
Successful End Condition	Sistem berhasil menampilkan menampilkan hasil pencarian berdasarkan kata kunci/*query* yang dimasukkan oleh pengguna	
Failed End Condition	Sistem tidak berhasil menampilkan menampilkan hasil pencarian berdasarkan kata kunci/*query* yang dimasukkan oleh pengguna	
Actors	Pengguna	
Trigger	Pengguna meminta sistem untuk melakukan proses pencarian	
Main Flow	**Step**	**Action**
	1.	Pilih Menu Aggregator
	2.	Sistem menampilkan halaman informasi semua berita
	3.	Masukkan kata kunci/*query* informasi berita yang ingin dicari
	4.	Sistem menampilkan hasil pencarian informasi berita

III.5.1.3 *Activity Diagram*

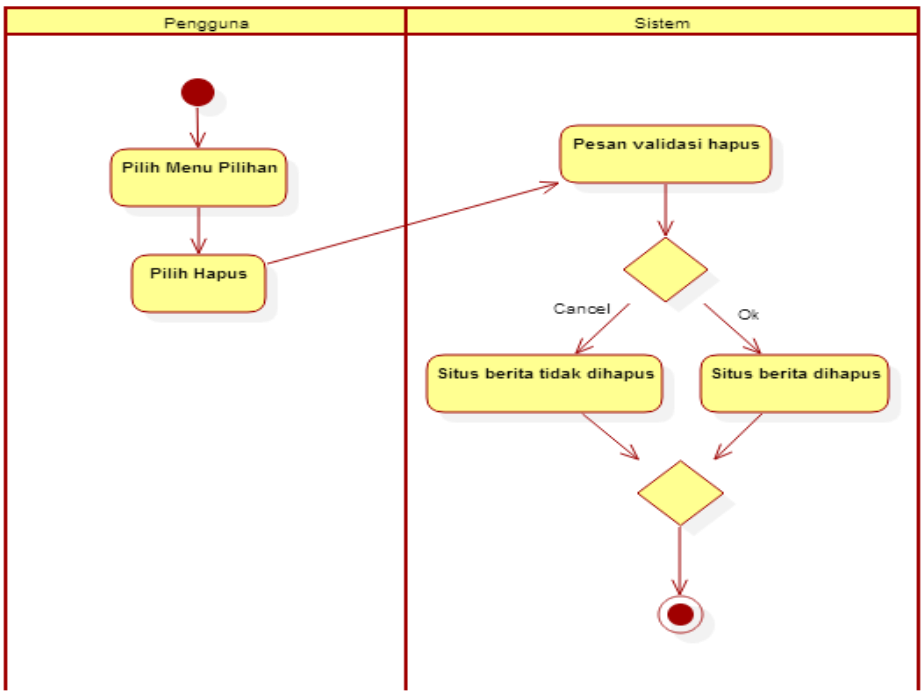
Activity diagram adalah teknik untuk menggambarkan logika prosedural, proses bisnis, dan jalur kerja. Berikut merupakan *activity diagram* dari pembangunan sistem ini berdasarkan dari *use case scenario* :



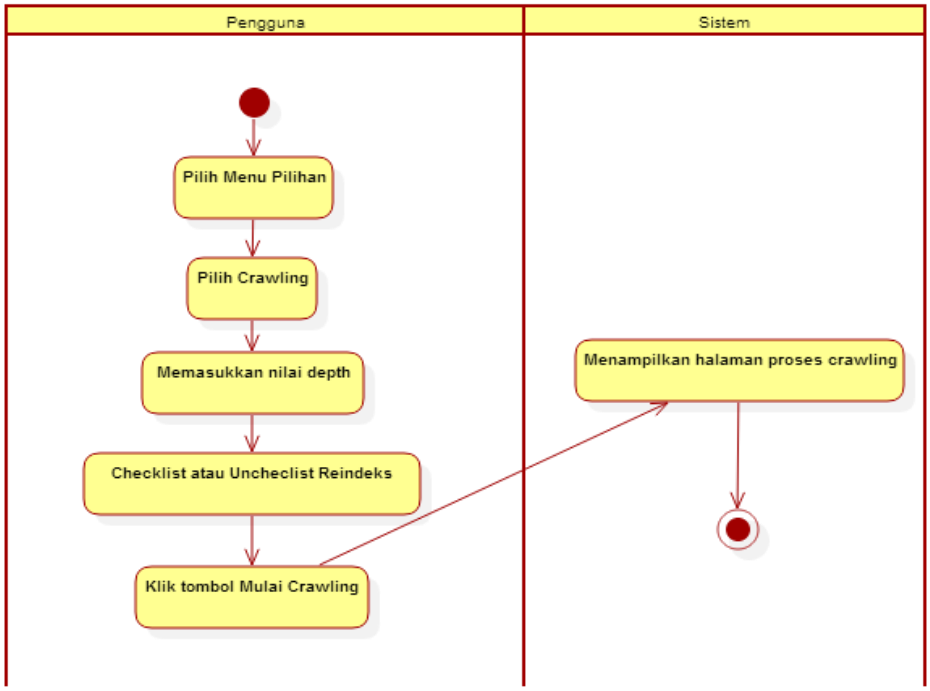
Gambar III.7 *Activity Diagram* Penyajian Tambah Situs Berita



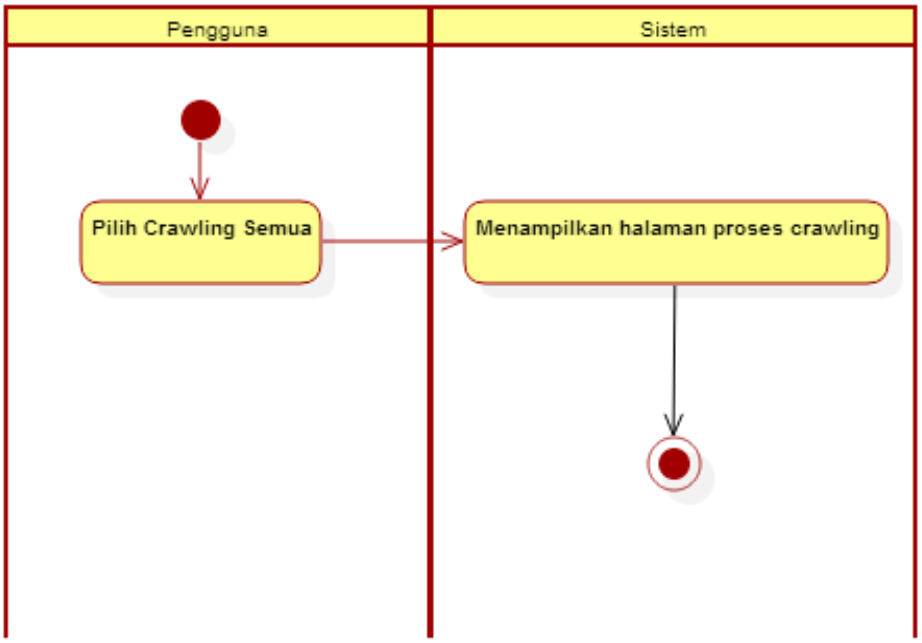
Gambar III.8 *Activity Diagram* Penyajian Edit Situs Berita



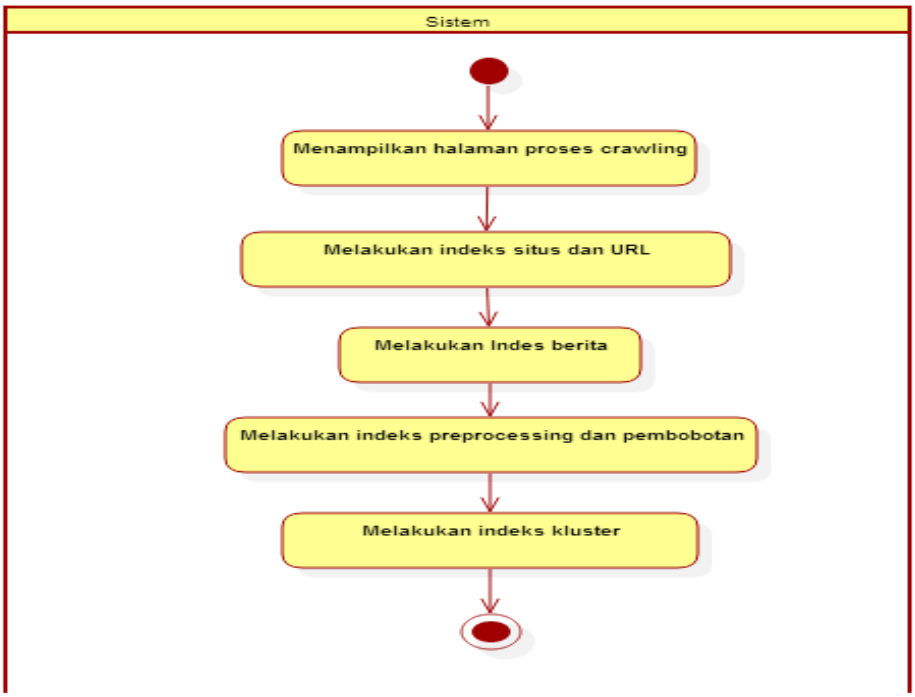
Gambar III.9 *Activity Diagram* Penyajian Hapus Situs Berita



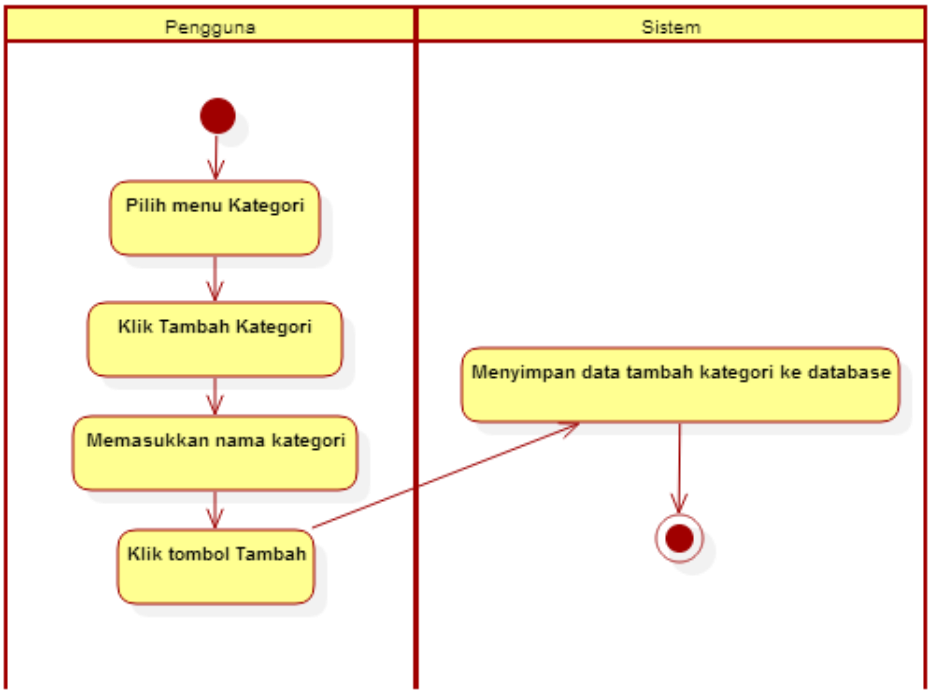
Gambar III.10 *Activity Diagram* Penyajian *Crawling* Situs Berita



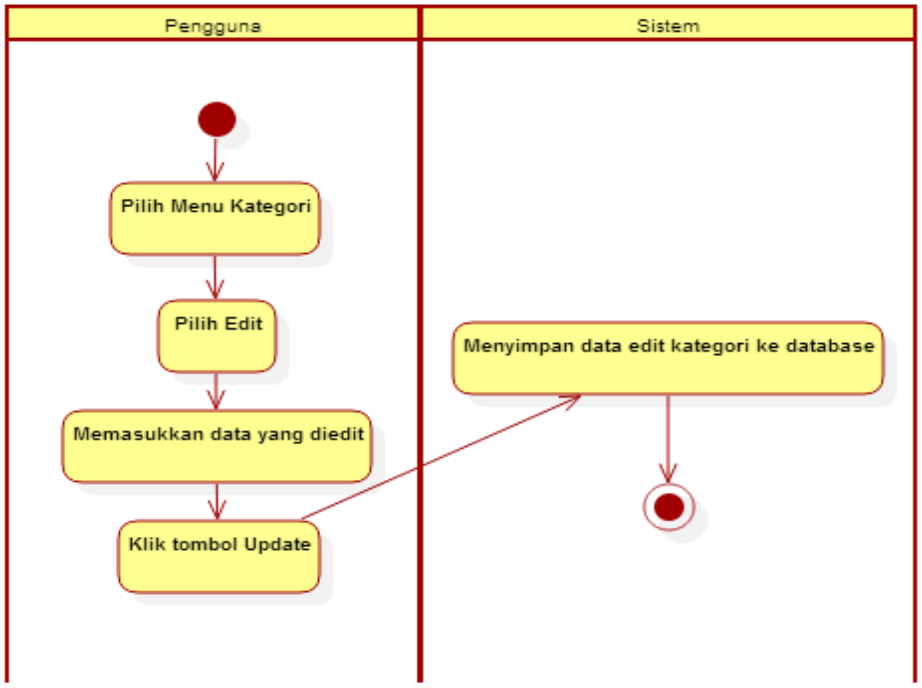
Gambar III.11 *Activity Diagram* Penyajian *Crawling* Semua Situs Berita



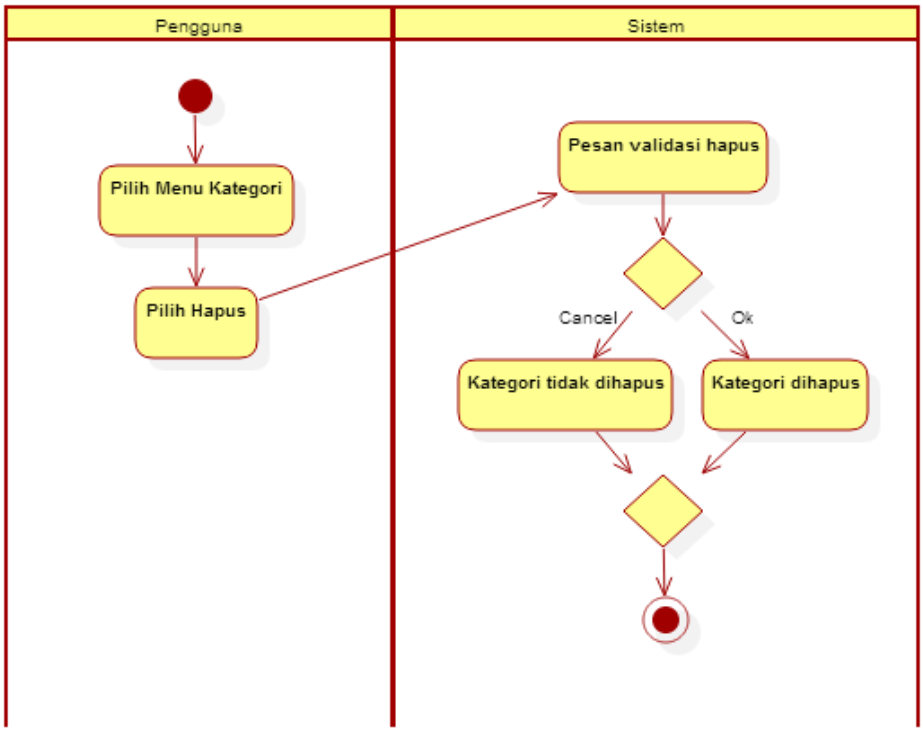
Gambar III.12 *Activity Diagram* Penyajian Hasil Proses *Crawling*



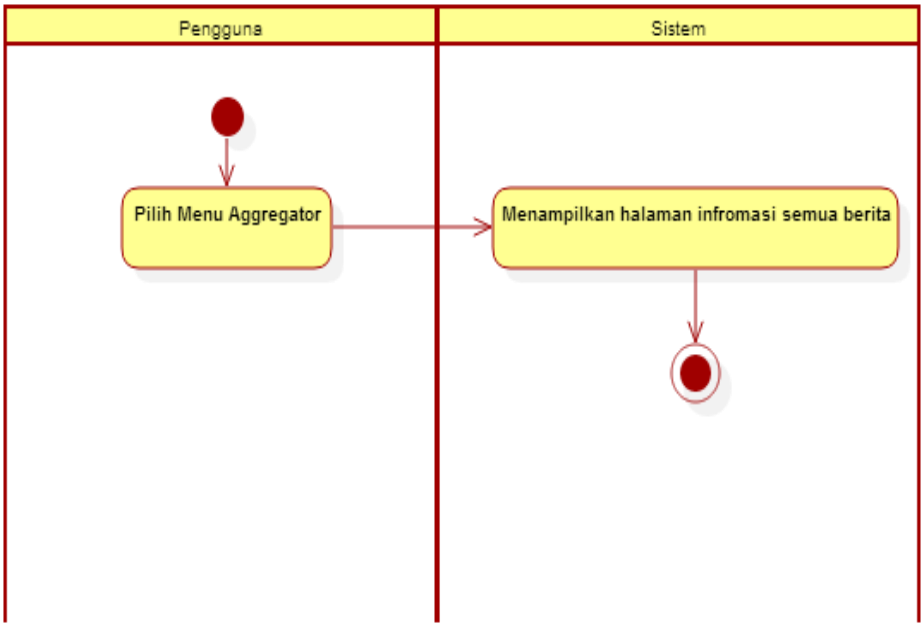
Gambar III.13 *Activity Diagram* Penyajian Tambah Kategori



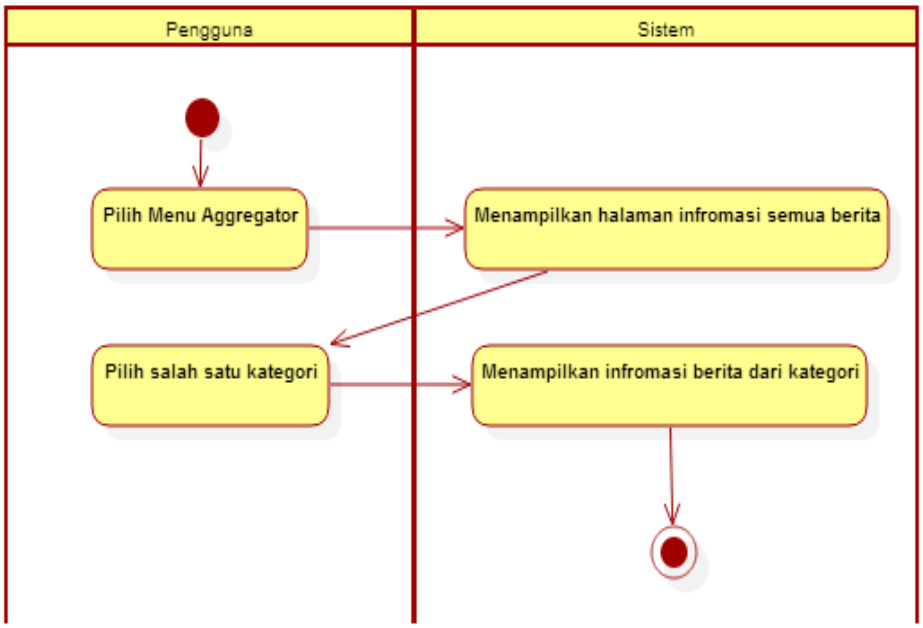
Gambar III.14 *Activity Diagram* Penyajian Edit Kategori



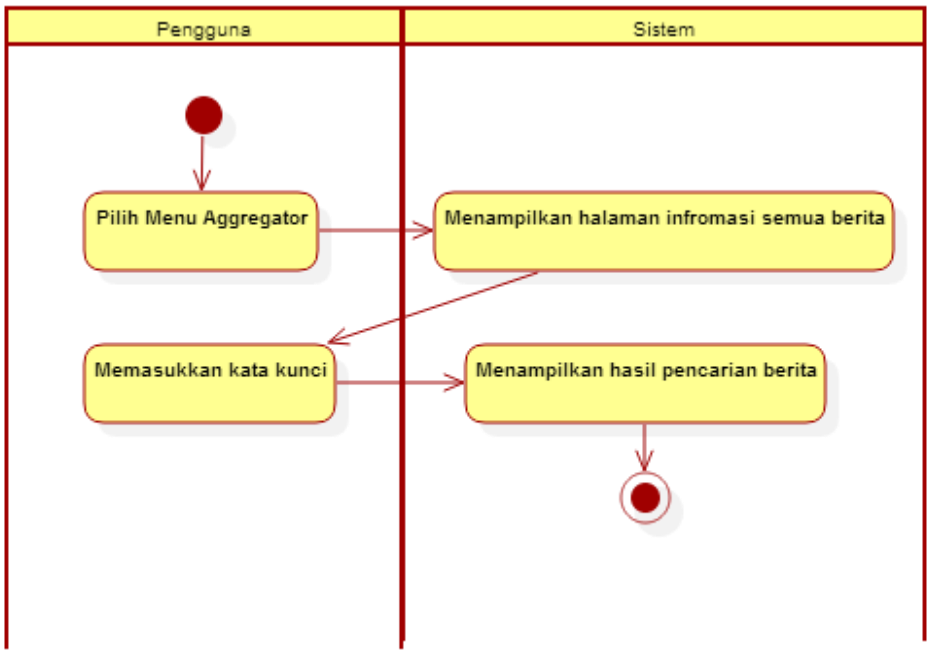
Gambar III.15 *Activity Diagram* Penyajian Hapus Kategori



Gambar III.16 *Activity Diagram* Penyajian Informasi Semua Berita



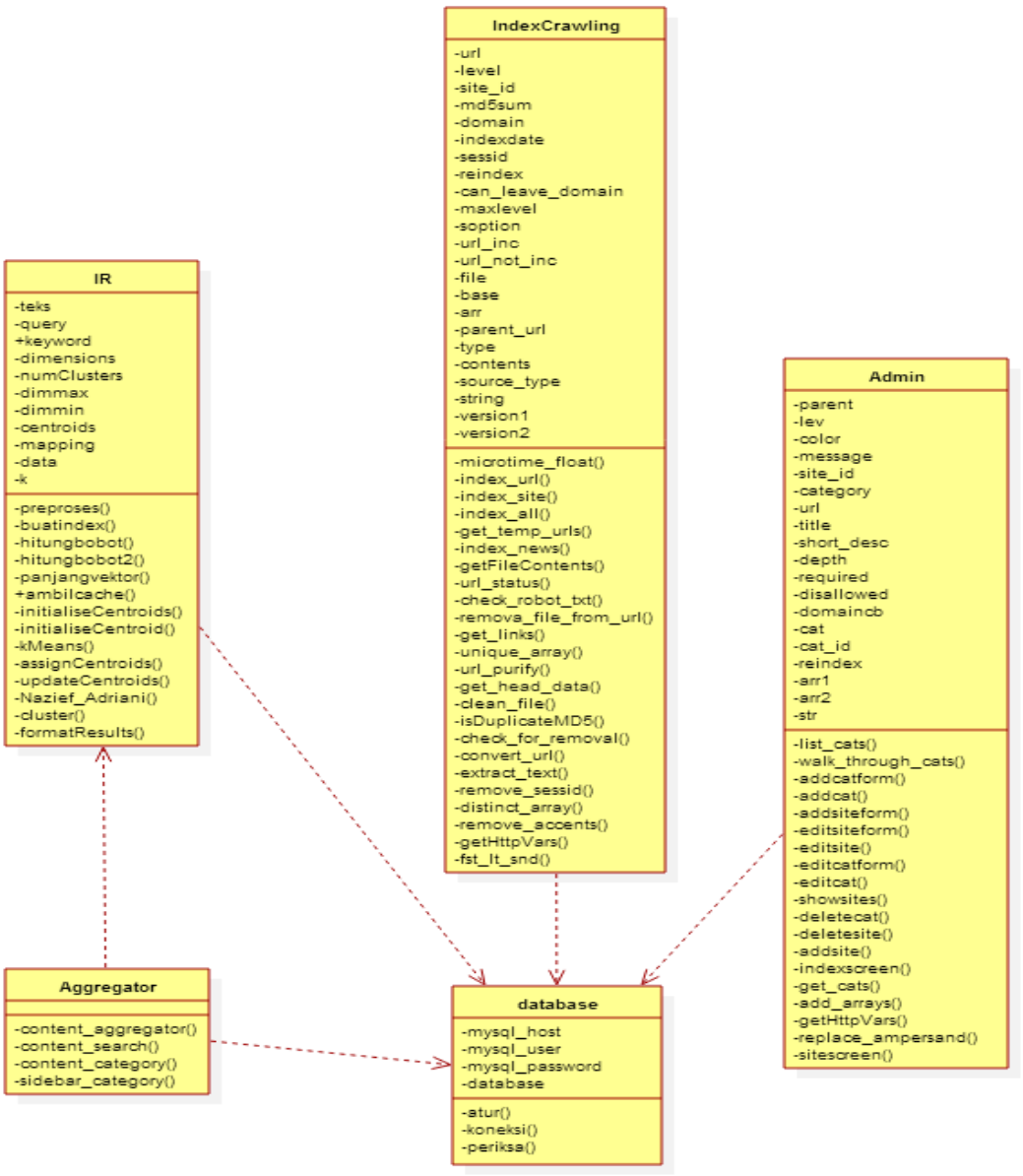
Gambar III.17 *Activity Diagram* Penyajian Informasi Berita Berdasarkan Kategori



Gambar III.18 *Activity Diagram* Penyajian Pencarian Informasi Berita

III.5.1.4 *Class Diagram*

Pada Gambar III.19 dibawah ini merupakan *class diagram* dari pembangunan sistem *news aggregator* ini.



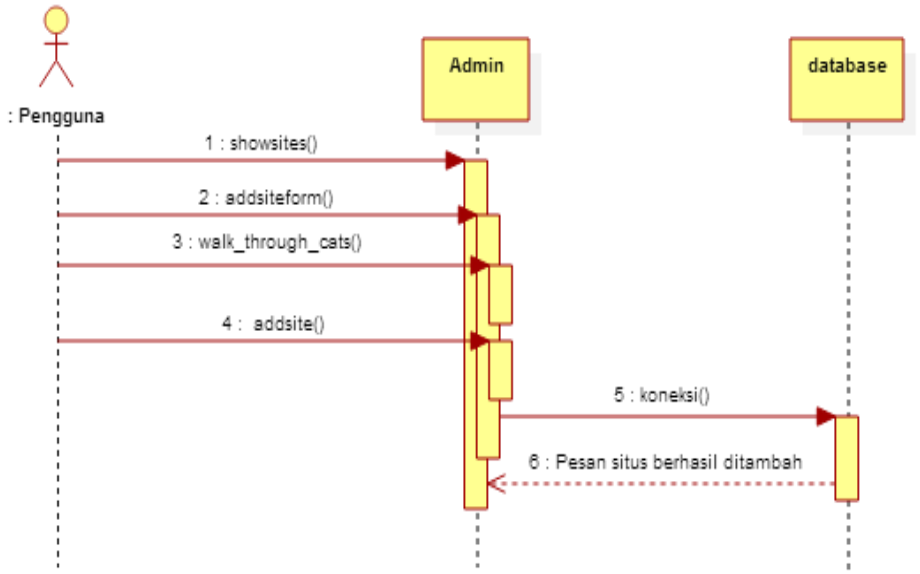
Gambar III.19 *Class Diagram* Sistem *News Aggregator*

III.5.1.5 *Sequence Diagram*

Pada bagian ini akan dipaparkan *sequence diagram* dari *usecase* yang berkaitan.

1. Penyajian Tambah Situs Berita

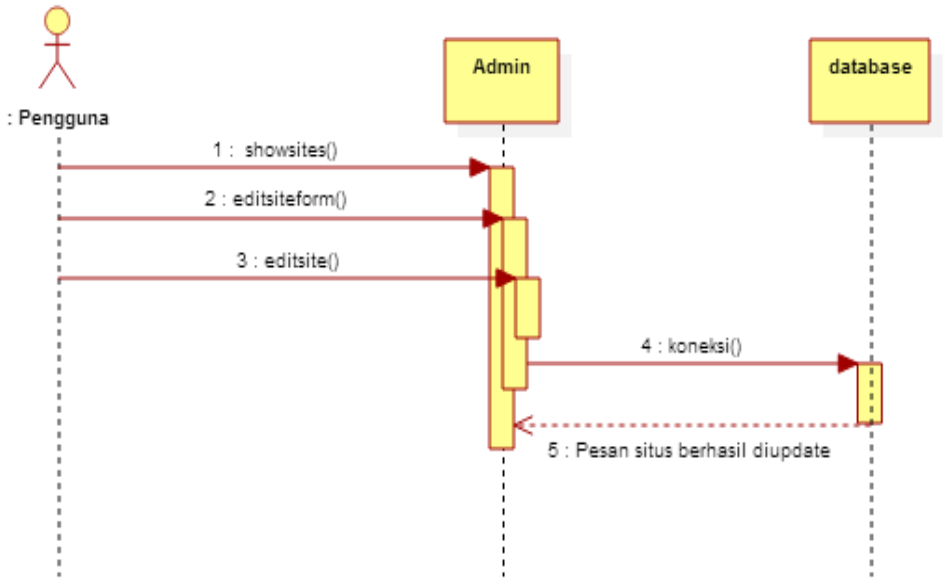
   Sequence diagram penyajian tambah situs berita dapat dilihat pada Gambar III.20 berikut:



Gambar III.20 *Sequence diagram* Penyajian Tambah Situs Berita

2. Penyajian Edit Situs Berita

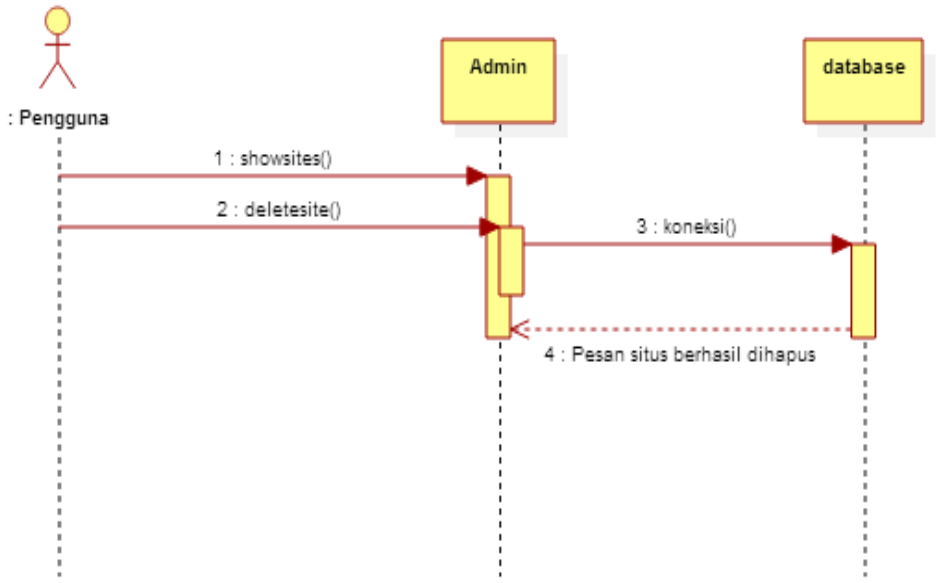
   Sequence diagram penyajian edit situs berita dapat dilihat pada Gambar III.21 berikut:



Gambar III.21 *Sequence diagram* Penyajian Edit Situs Berita

3. Penyajian Hapus Situs Berita

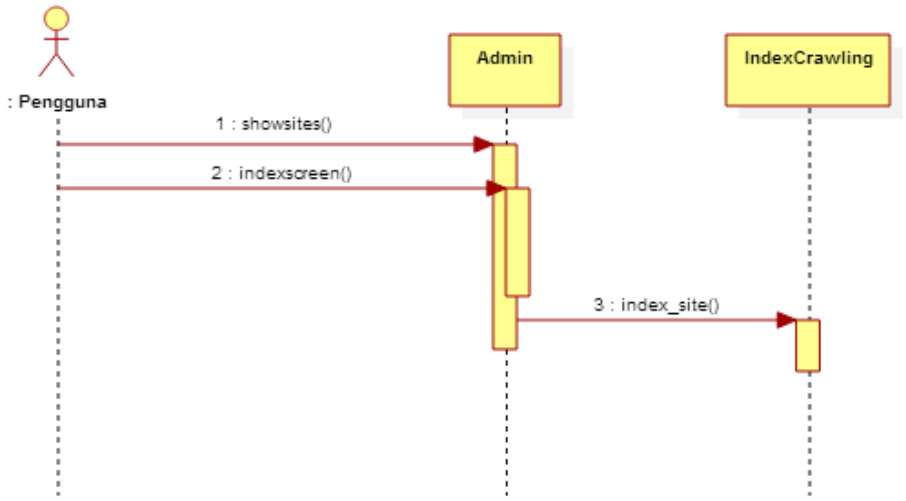
Sequence diagram penyajian hapus situs berita dapat dilihat pada Gambar III.22 berikut:



Gambar III.22 *Sequence diagram* Penyajian Hapus Situs Berita

4. Penyajian *Crawling* Situs Berita

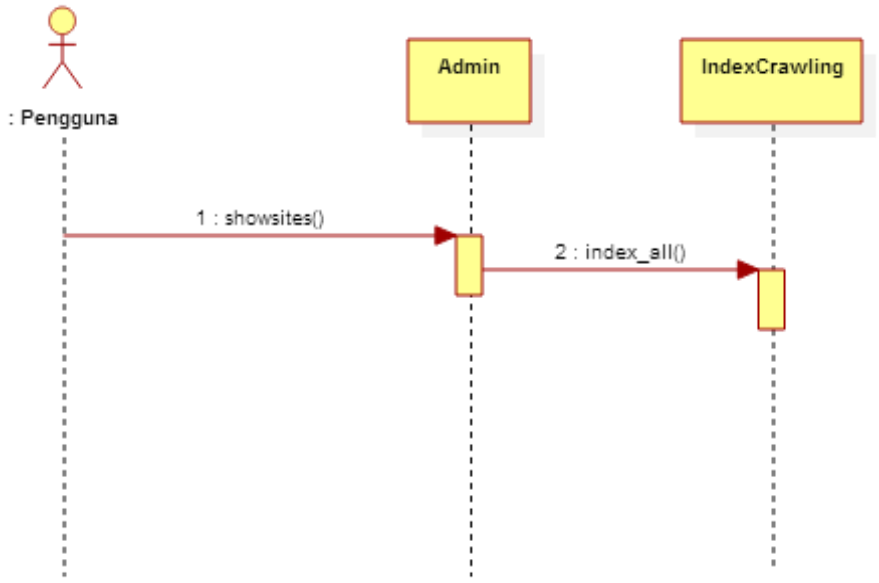
   Sequence diagram penyajian *crawling* situs berita dapat dilihat pada Gambar III.23 berikut:



Gambar III.23 *Sequence diagram* Penyajian *Crawling* Situs Berita

5. Penyajian *Crawling* Semua Situs Berita

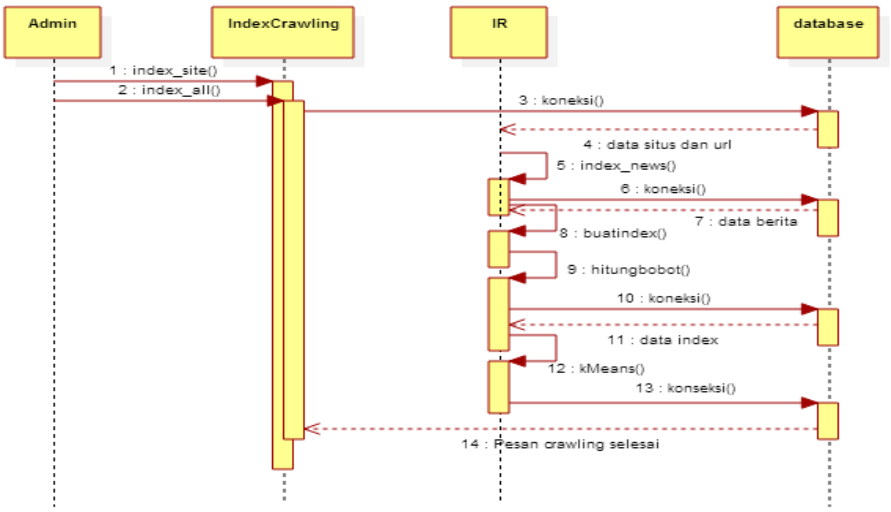
Sequence diagram penyajian *crawling* semua situs berita dapat dilihat pada Gambar III.24 berikut:



Gambar III.24 *Sequence diagram* Penyajian *Crawling* Semua Situs Berita

6. Penyajian Hasil Proses *Crawling*

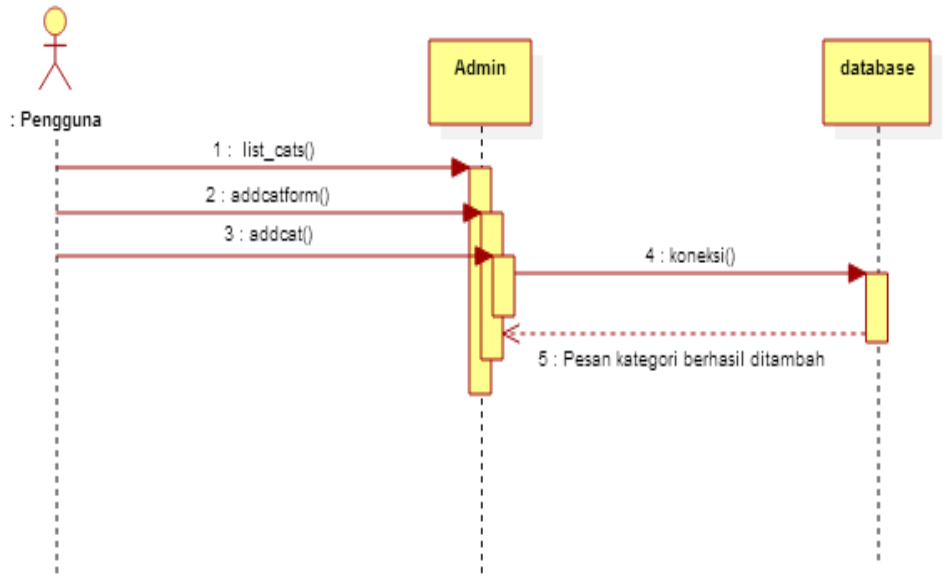
   Sequence diagram penyajian hasil proses *crawling* dapat dilihat pada Gambar III.25 berikut:



Gambar III.25 *Sequence diagram* Penyajian Hasil Proses *Crawling*

7. Penyajian Tambah Kategori

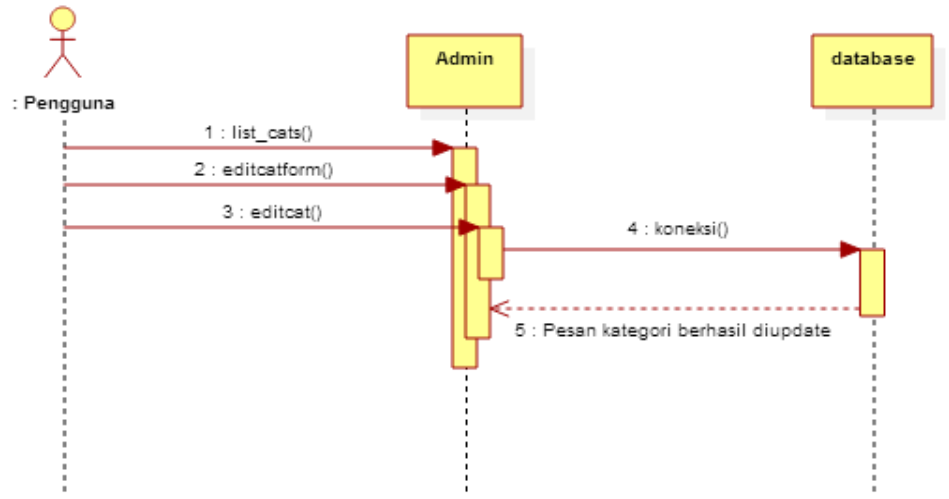
Sequence diagram penyajian tambah kategori dapat dilihat pada Gambar III.26 berikut:



Gambar III.26 *Sequence diagram* Penyajian Tambah Kategori

8. Penyajian Edit Kategori

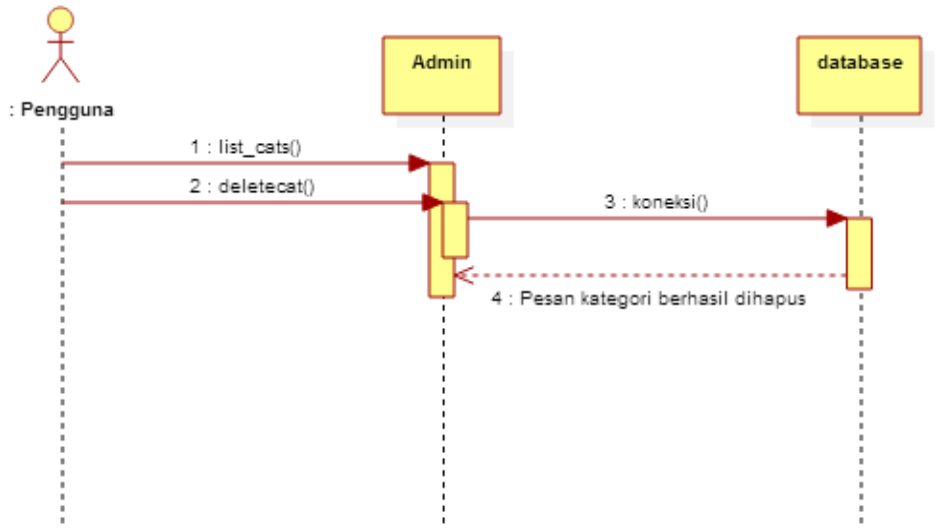
   Sequence diagram penyajian edit kategori dapat dilihat pada Gambar III.27 berikut:



Gambar III.27 *Sequence diagram* Penyajian Edit Kategori

9. Penyajian Hapus Kategori

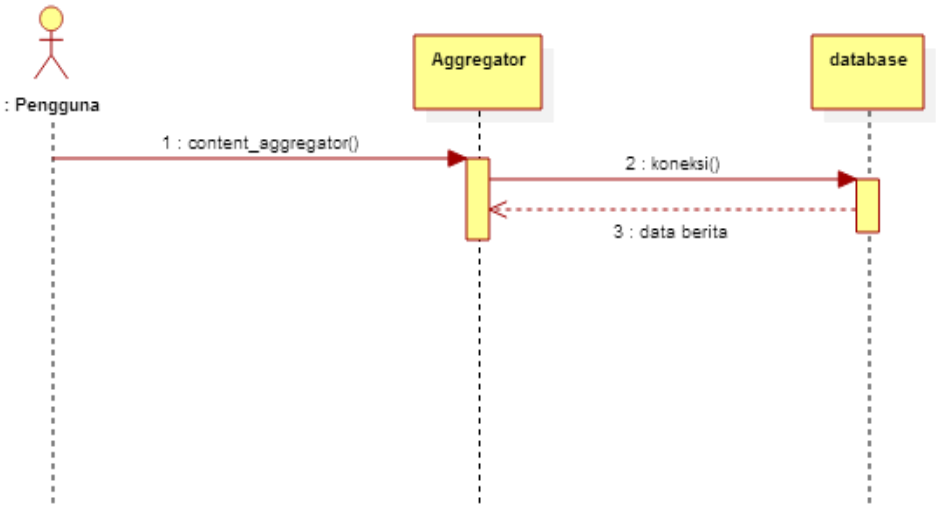
Sequence diagram penyajian hapus kategori dapat dilihat pada Gambar III.28 berikut:



Gambar III.28 *Sequence diagram* Penyajian Hapus Kategori

10. Penyajian Informasi Semua Berita

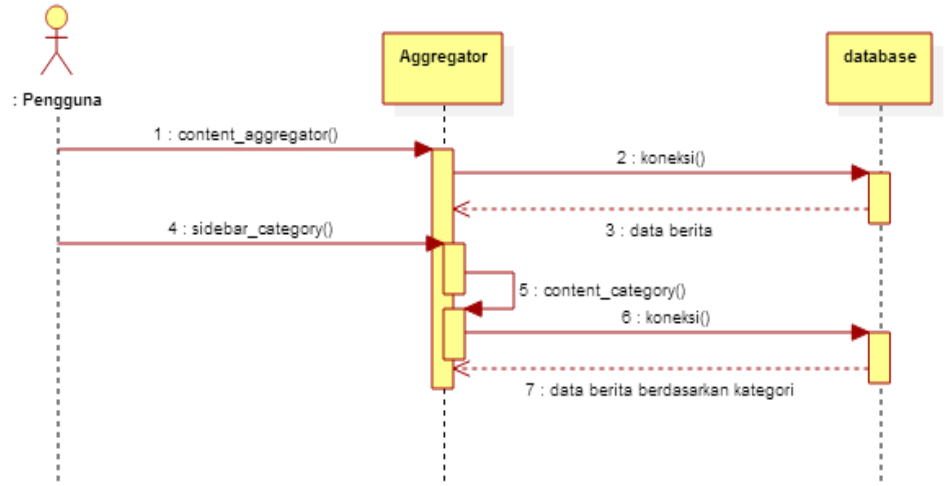
Sequence diagram penyajian informasi semua berita dapat dilihat pada Gambar III.29 berikut:



Gambar III.29 *Sequence diagram* Penyajian Informasi Semua Berita

11. Penyajian Informasi Berita Berdasarkan Kategori

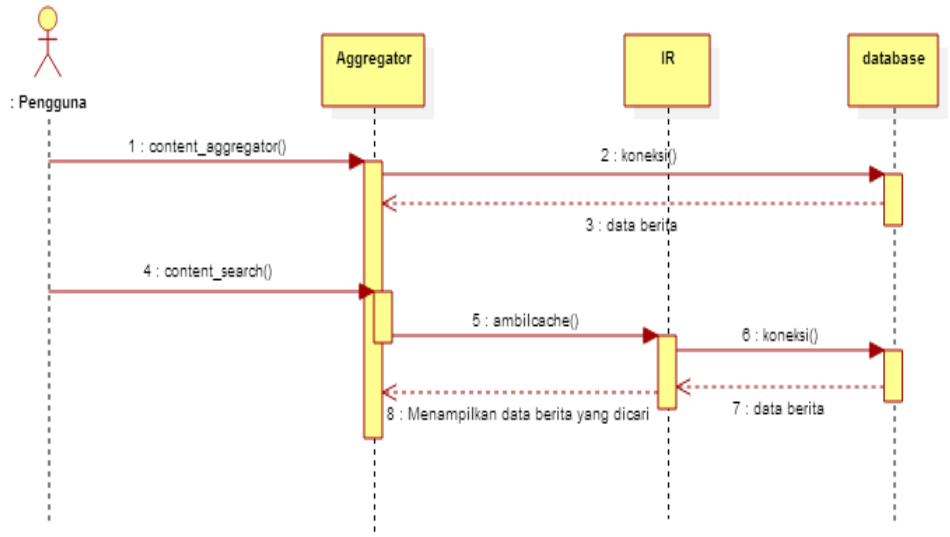
Sequence diagram penyajian informasi berita berdasarkan kategori dapat dilihat pada Gambar III.30 berikut:



Gambar III.30 *Sequence diagram* Penyajian Informasi Berita Berdasarkan Kategori

12. Penyajian Pencarian Informasi Berita

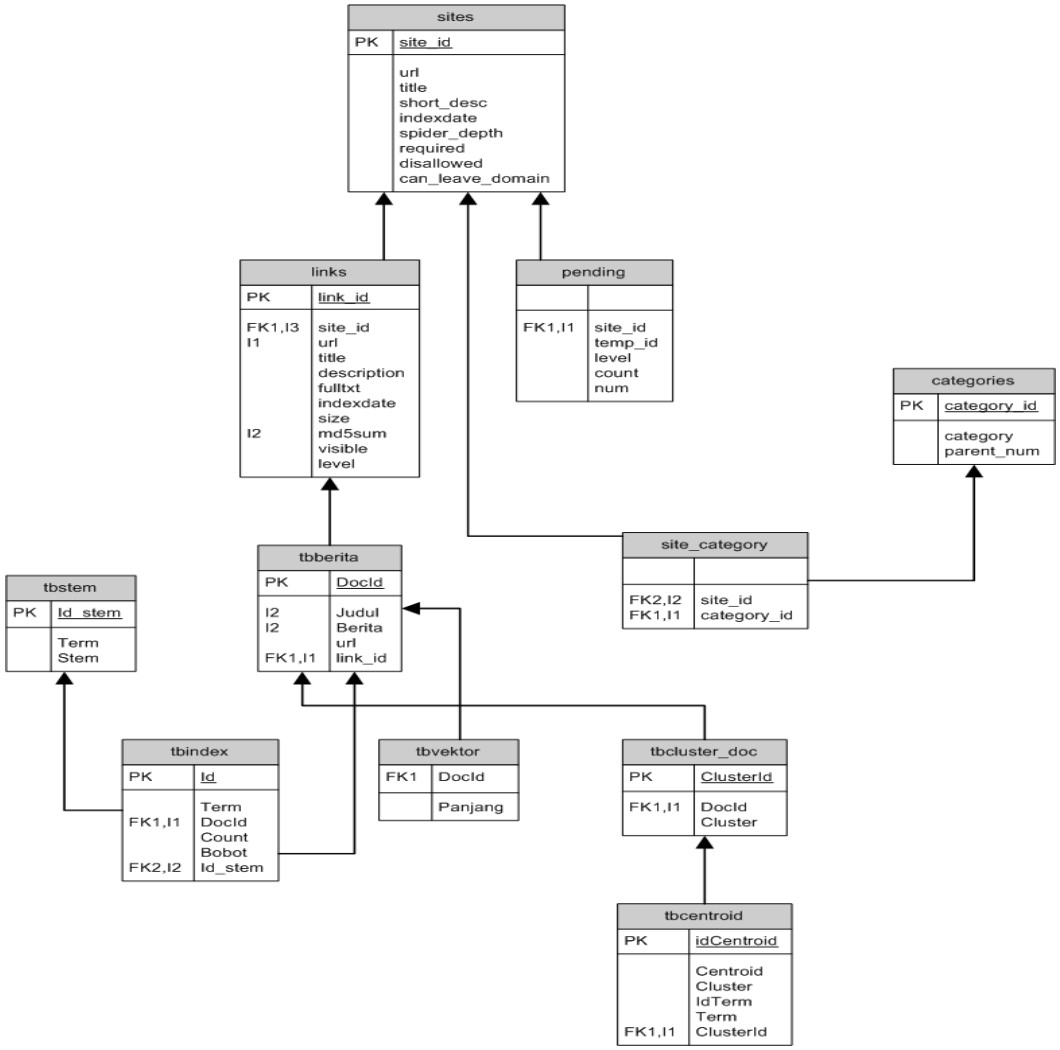
Sequence diagram penyajian pencarian informasi berita dapat dilihat pada berikut:



Gambar III.31 *Sequence diagram* Penyajian Pencarian Informasi Berita

III.6 Perancangan Data

Perancangan data dari pembangunan sistem *news aggregator* ini bisa dilihat pada skema relasi yang ditunjukkan pada Gambar III.32.



Gambar III.32 Skema Relasi Sistem *News Aggregator*

Berdasarkan pada Gambar III.32, maka dibuat perancangan tabel yang akan digunakan dalam sistem ini yang sebagai berikut :

Tabel III.29 Tabel categories

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
category_id	int	11	NO	Primary	auto_increment
category	text	-	YES		
parent_num	int	11	YES		

Tabel III.30 Tabel links

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
link_id	int	11	NO	Primary	auto_increment
site_id	int	11	YES	Foreign	
url	varchar	255	NO		
title	varchar	200	YES		
description	varchar	255	YES		
fulltxt	mediumtext	-	YES		
indexdate	date	-	YES		
size	float	-	YES		
md5sum	varchar	32	YES		
visible	int	11	YES		
level	int	11	YES		

Tabel III.31 Tabel pending

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
site_id	int	11	YES	Foreign	
temp_id	varchar	32	YES		
level	int	11	YES		
count	int	11	YES		
num	int	11	YES		

Tabel III.32 Tabel site_category

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
site_id	int	11	YES	Foreign	
category_id	int	11	YES	Foreign	

Tabel III.33 Tabel sites

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
site_id	int	11	NO	Primary	
url	varchar	255	YES		
title	varchar	255	YES		
short_desc	text	-	YES		
indexdate	date	-	YES		
spider_depth	int	11	YES		
required	text	-	YES		
disallowed	text	-	YES		
can_leave_domain	tinyint	1	YES		

Tabel III.34 Tabel tbberita

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
DocId	int	11	NO	Primary	auto_increment
Judul	varchar	100	NO		
Berita	varchar	255	NO		
url	varchar	255	NO		
link_id	int	11	YES	Foreign	

Tabel III.35 Tabel tbcluster_doc

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
ClusterId	int	11	NO	Primary	auto_increment
DocId	int	11	NO	Foreign	
Cluster	int	11	NO		

Tabel III.36 Tabel tbcentroid

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
IdCentroid	int	11	NO	Primary	auto_increment
Centroid	float	-	NO		
Cluster	int	11	NO		
IdTerm	int	11	NO		
Term	varchar	30	NO		
ClusterId	int	11	YES	Foreign	

Tabel III.37 Tabel tbindex

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
Id	int	11	NO	Primary	auto_increment
Term	varchar	30	NO		
DocId	int	11	NO	Foreign	
Count	int	11	NO		
Bobot	float	-	NO		
Id_stem	int	11	YES	Foreign	

Tabel III.38 Tabel tbstem

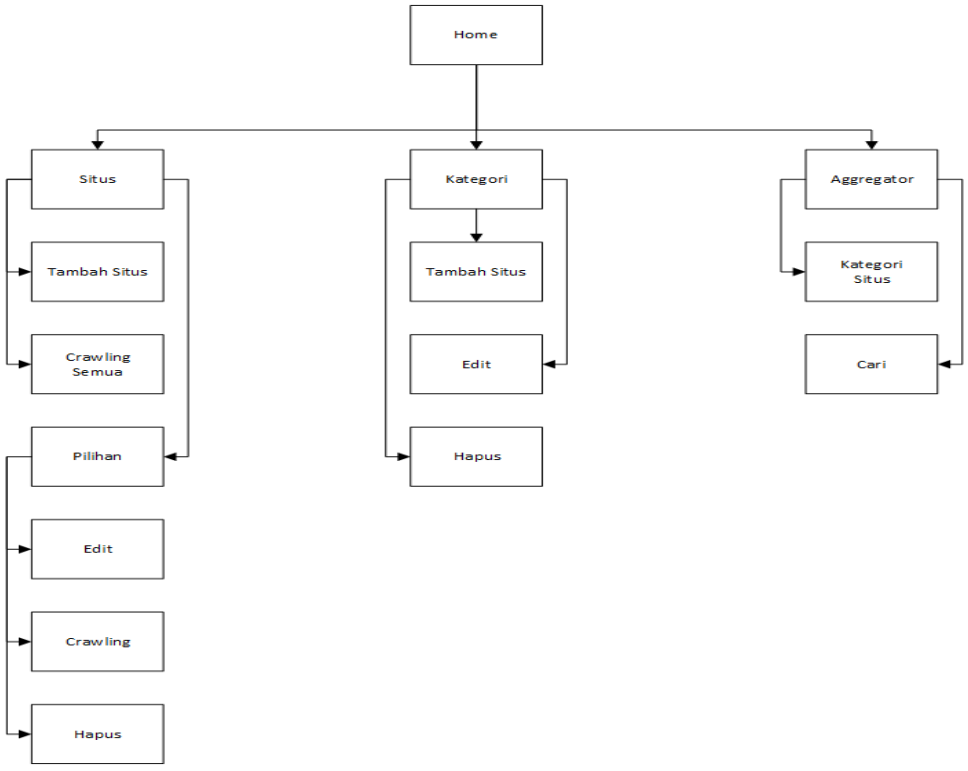
Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
Id_stem	int	11	NO	Primary	auto_increment
Term	varchar	30	NO		
Stem	varchar	30	NO		

Tabel III.39 Tabel tbvektor

Nama *Field*	Tipe Data	Panjang	*Null*	Kunci	*Extra*
DocId	int	11	NO	Foreign	
Panjang	float	-	NO		

III.7 Perancangan Struktur Menu

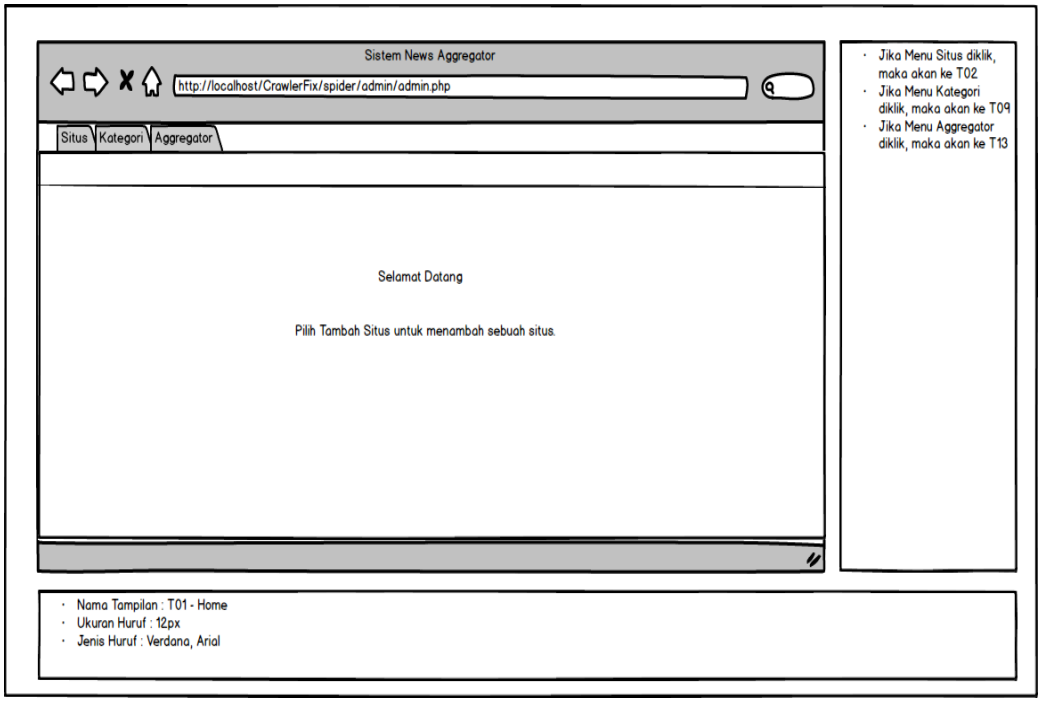
Perancangan struktur menu merupakan gambaran jalur pemakaian aplikasi sehingga aplikasi yang dibangun mudah dipahami dan mudah digunakan. Perancangan struktur menu dari sistem ini dapa dilihat pada Gambar III.33.



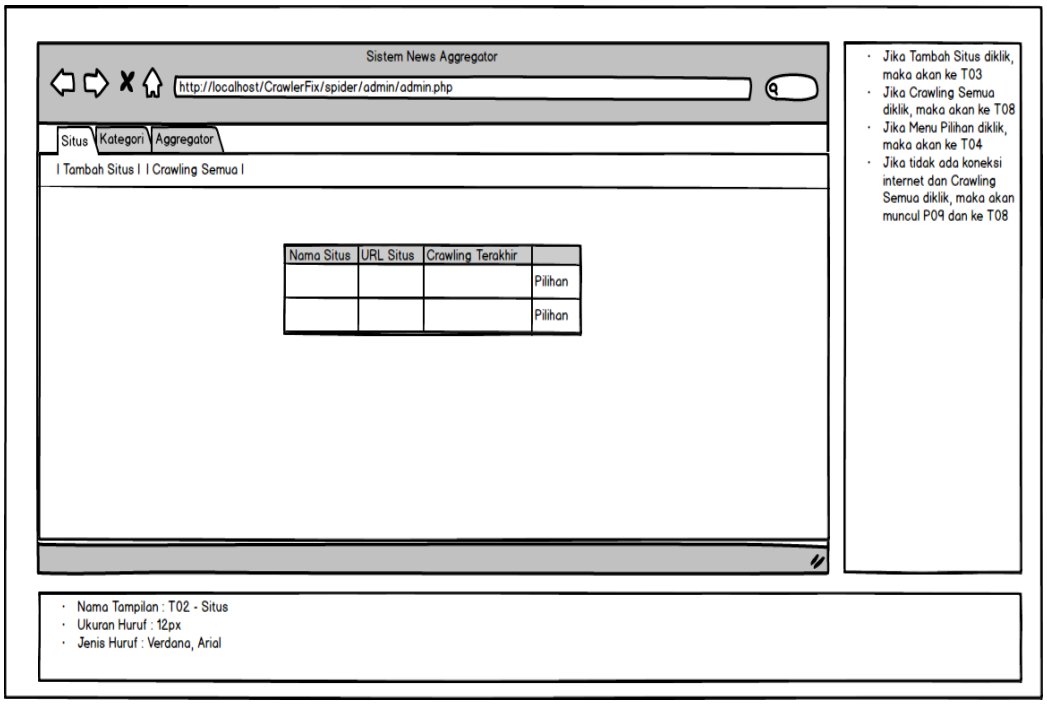
Gambar III.33 Struktur Menu

III.8 Perancangan Antarmuka

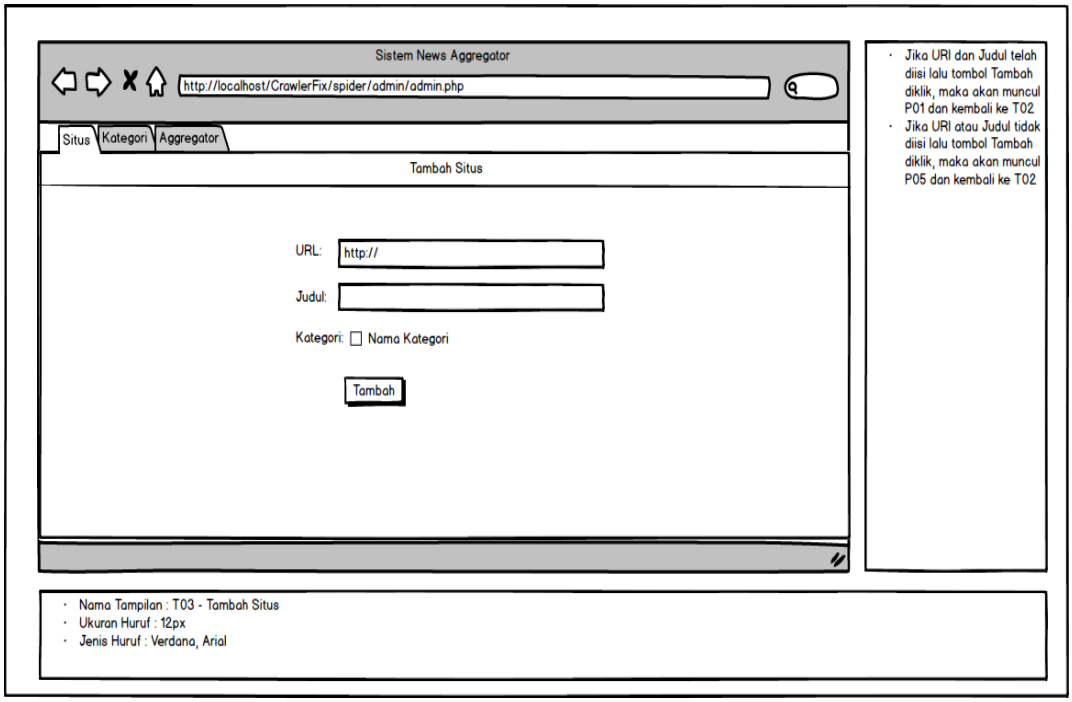
Perancangan antarmuka mendeskripsikan rencana tampilan dari setiap *form* yang akan diimplementasikan pada sistem *news aggregator*. Berikut adalah perancangan antarmuka sistem *news aggregator*.



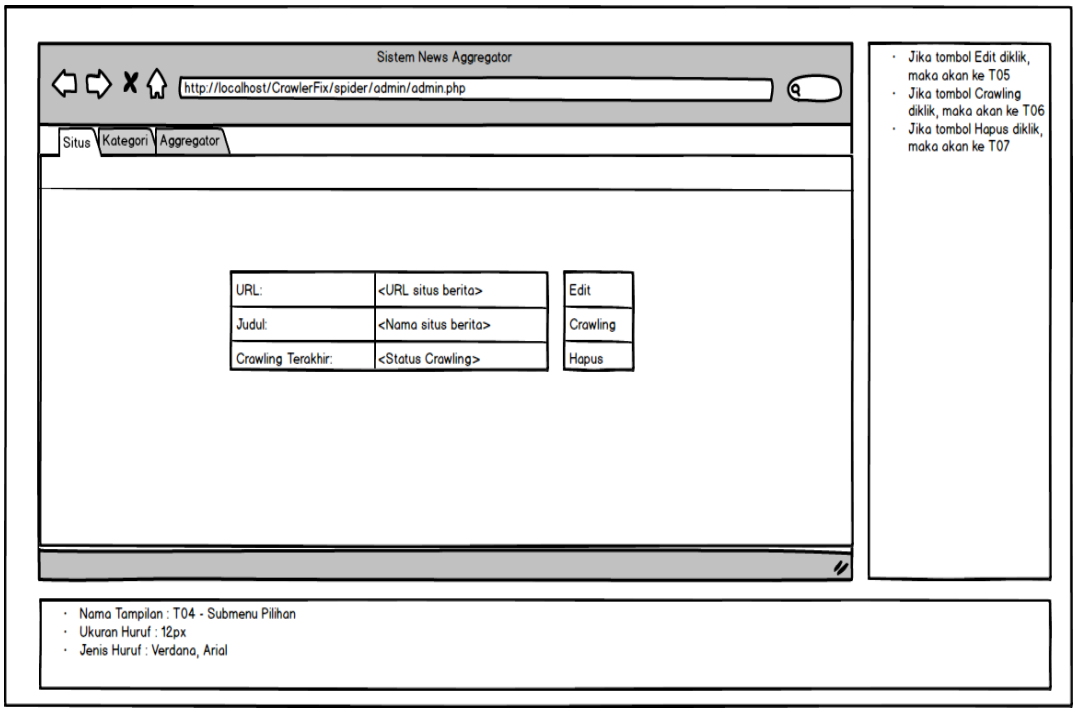
Gambar III.34 Perancangan Antarmuka T01 – Home



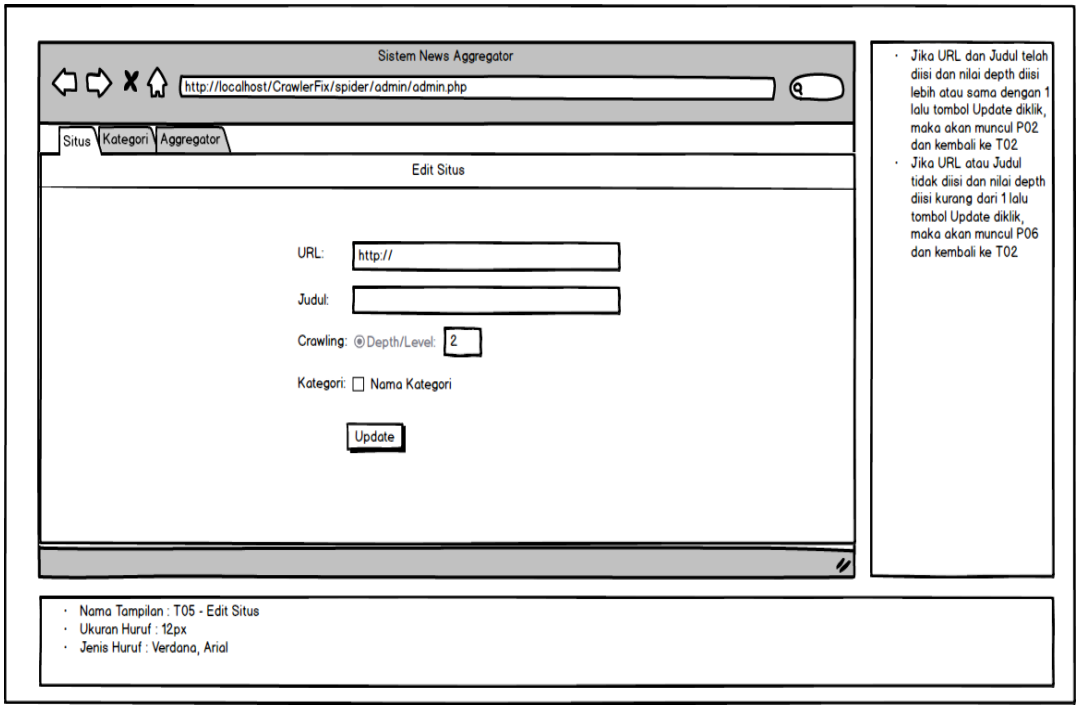
Gambar III.35 Perancangan Antarmuka T02 – Situs



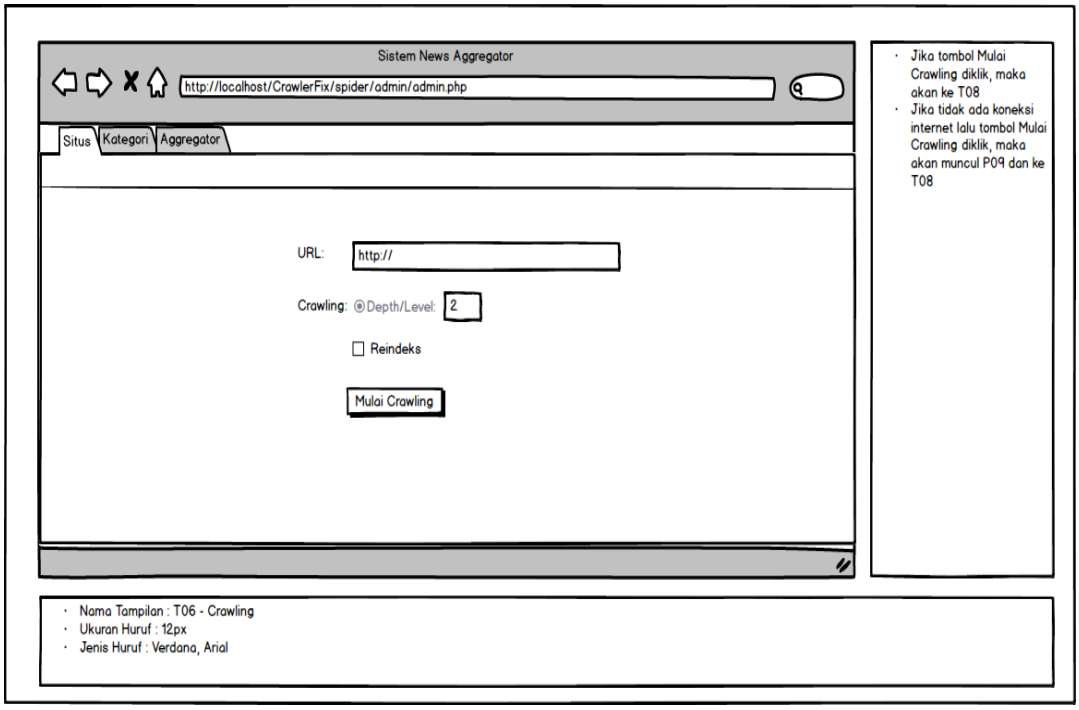
Gambar III.36 Perancangan Antarmuka T03 – Tambah Situs



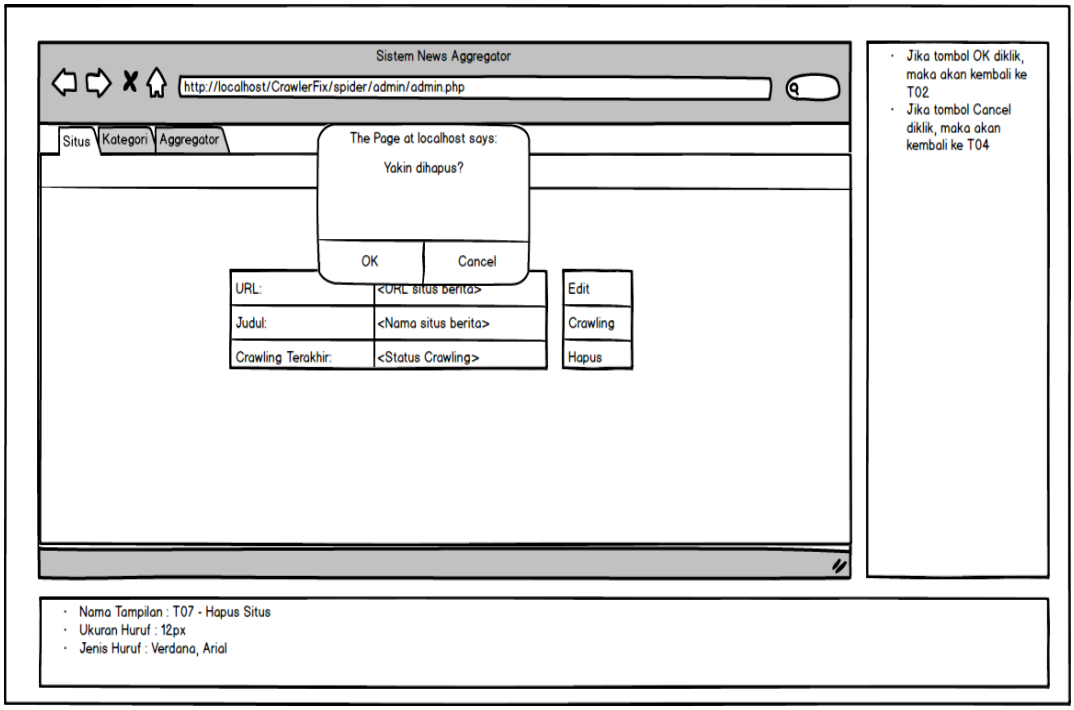
Gambar III.37 Perancangan Antarmuka T04 – Submenu Pilihan



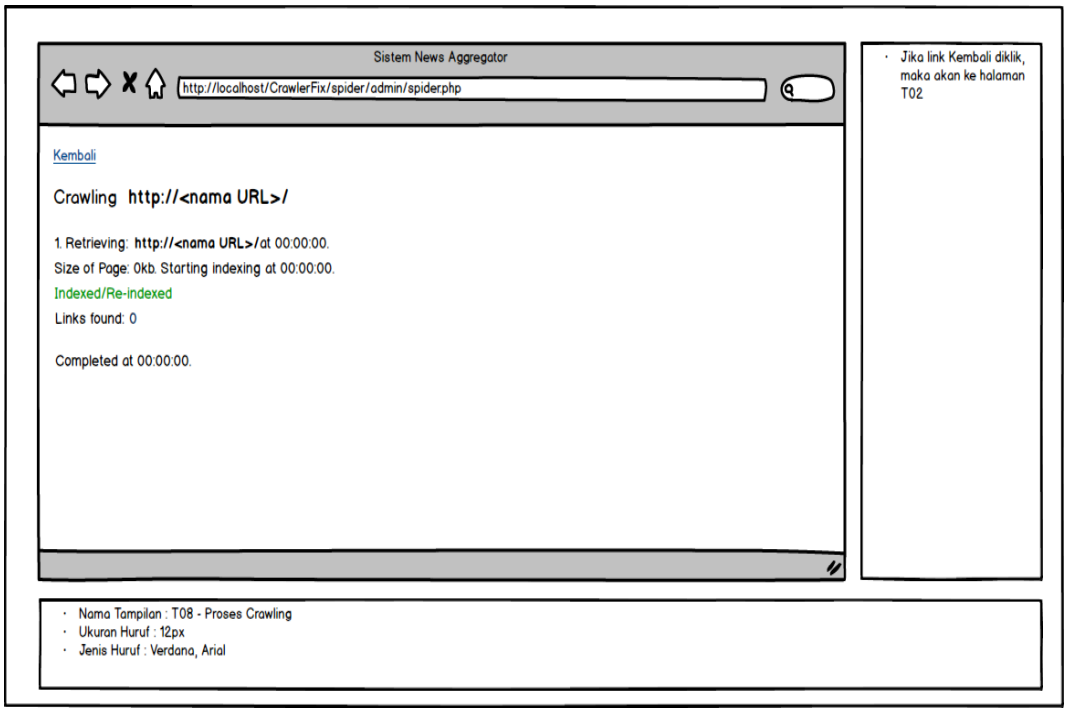
Gambar III.38 Perancangan Antarmuka T05 – Edit Situs



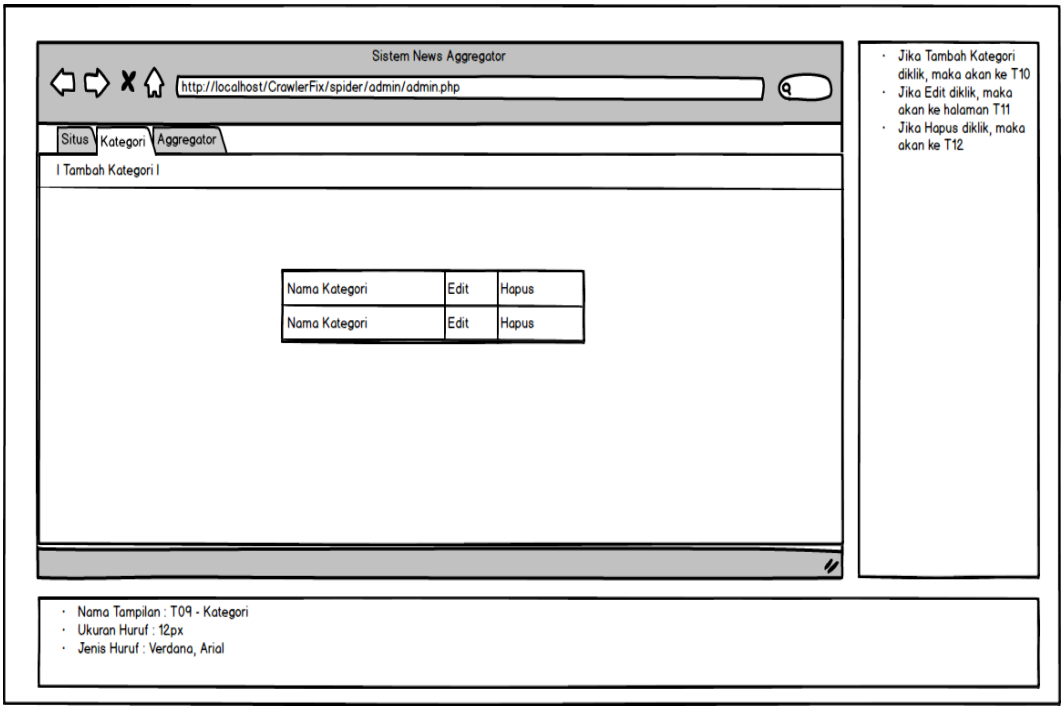
Gambar III.39 Perancangan Antarmuka T06 – Crawling



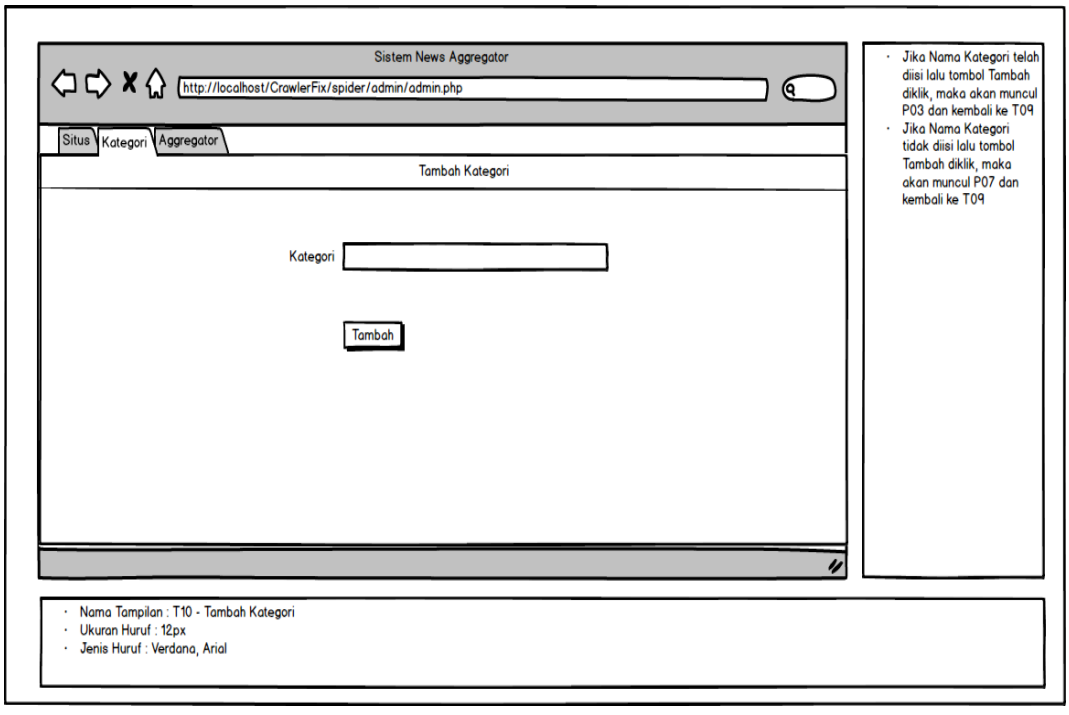
Gambar III.40 Perancangan Antarmuka T07 – Hapus Situs



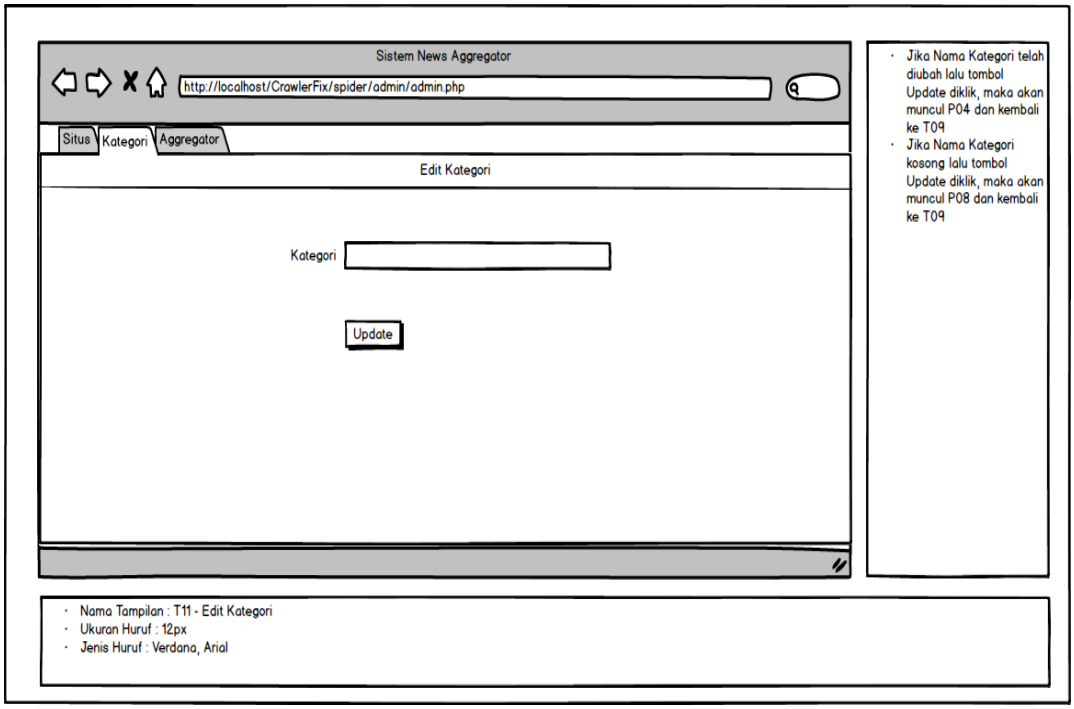
Gambar III.41 Perancangan Antarmuka T08 – Proses Crawling



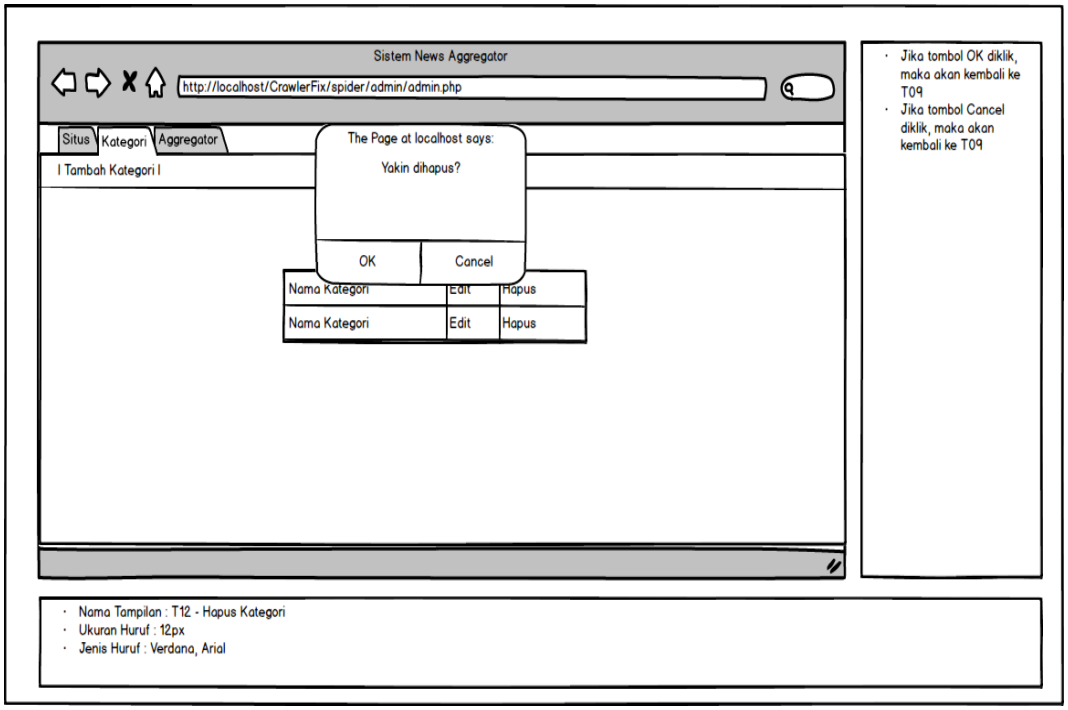
Gambar III.42 Perancangan Antarmuka T09 – Kategori



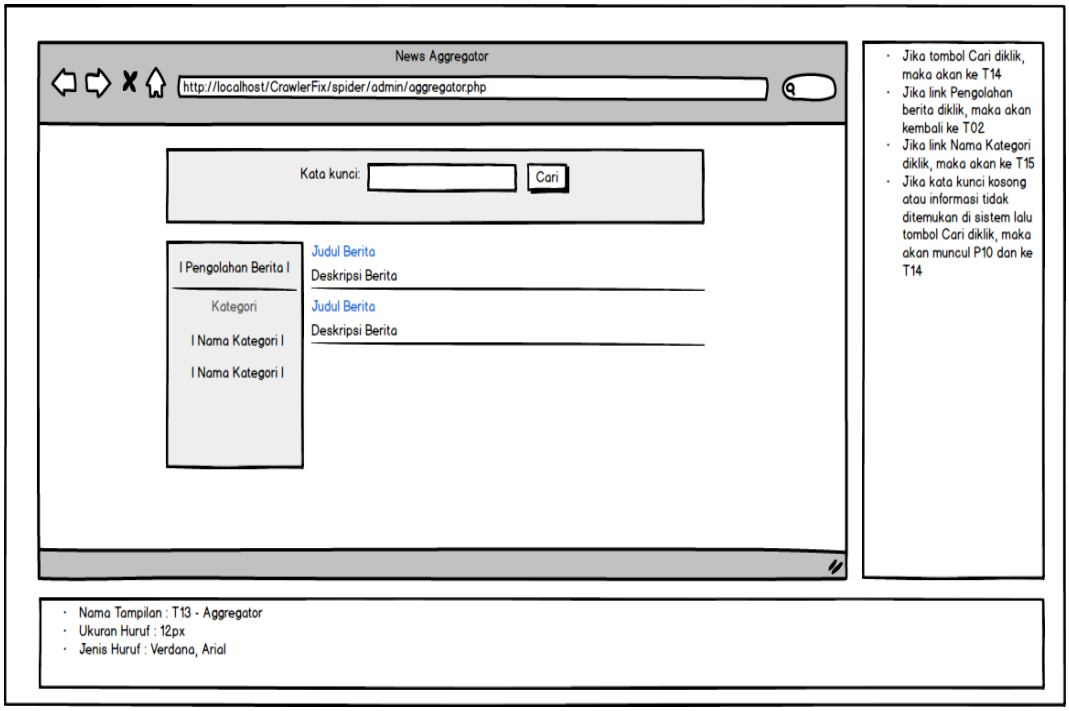
Gambar III.43 Perancangan Antarmuka T10 – Tambah Kategori



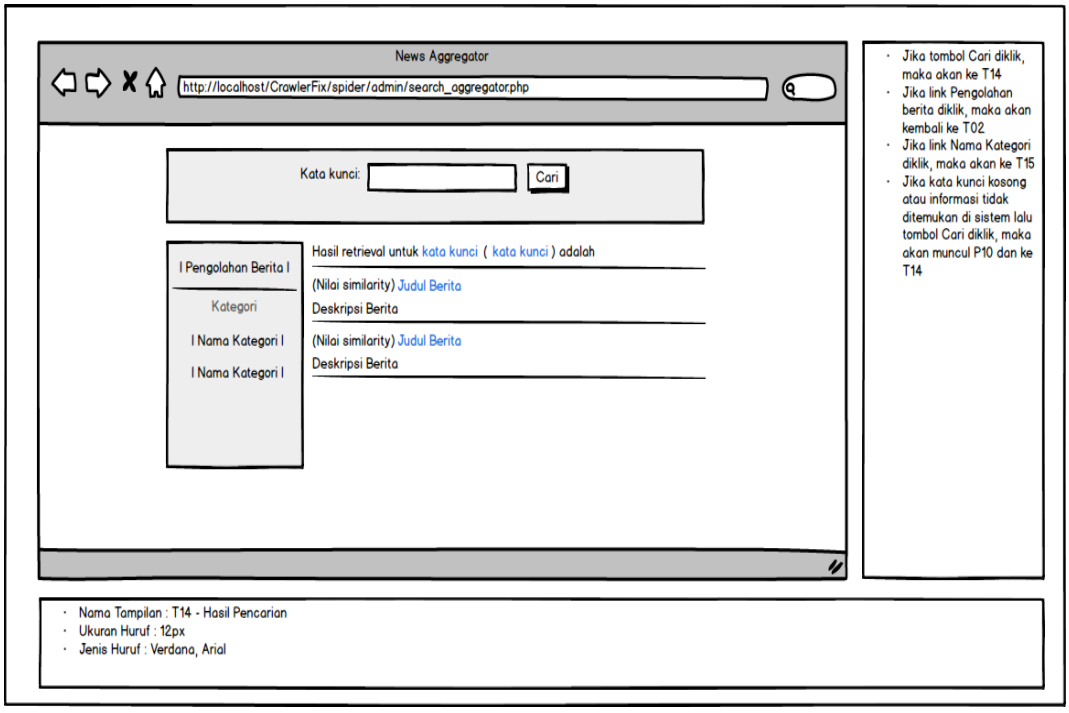
Gambar III.44 Perancangan Antarmuka T11 – Edit Kategori



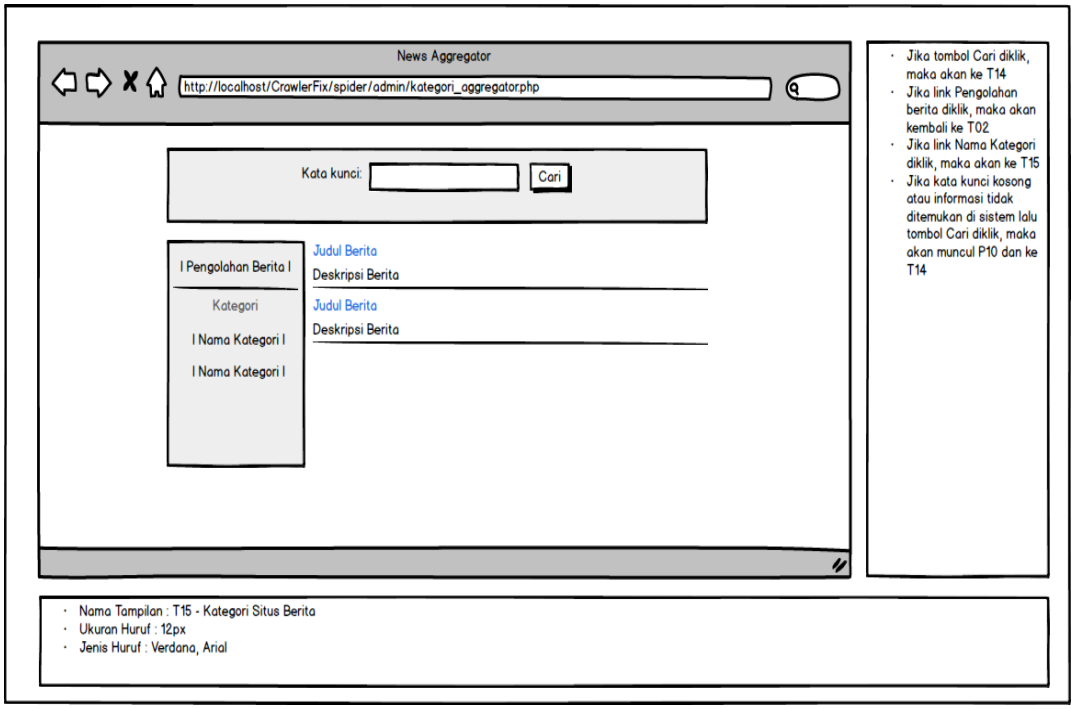
Gambar III.45 Perancangan Antarmuka T12 – Hapus Kategori



Gambar III.46 Perancangan Antarmuka T13 – Aggregator



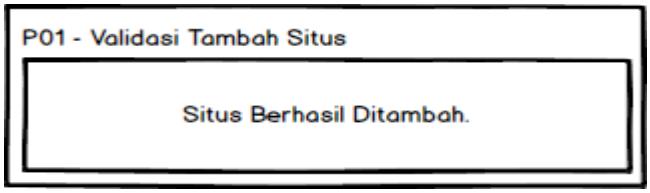
Gambar III.47 Perancangan Antarmuka T14 – Hasil Pencarian



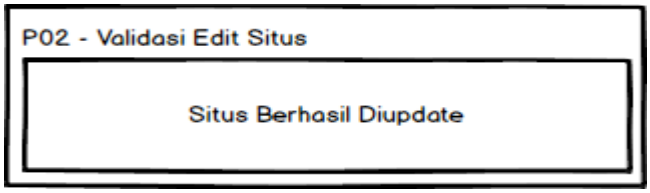
Gambar III.48 Perancangan Antarmuka T15 – Kategori Situs Berita

III.9 Perancangan Pesan

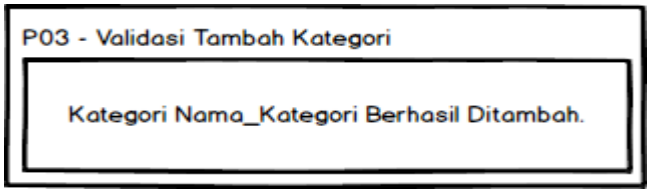
Perancangan pesan mendeskripsikan rencana tampilan dari setiap pesan yang muncul dalam sistem *news aggregator*. Berikut adalah perancangan pesan sistem *news aggregator*.



Gambar III.49 Perancangan Pesan P01 – Validasi Tambah Situs



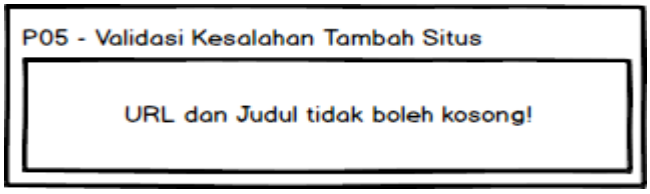
Gambar III.50 Perancangan Pesan P02 – Validasi Edit Situs



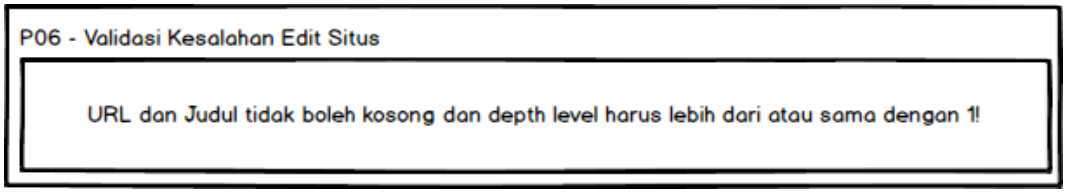
Gambar III.51 Perancangan Pesan P03 – Validasi Tambah Kategori



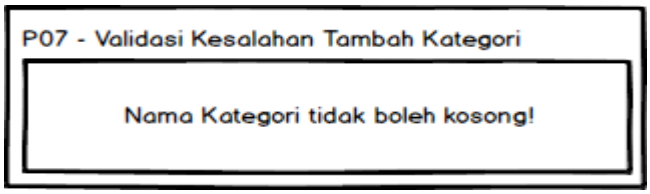
Gambar III.52 Perancangan Pesan P04 – Validasi Edit Kategori



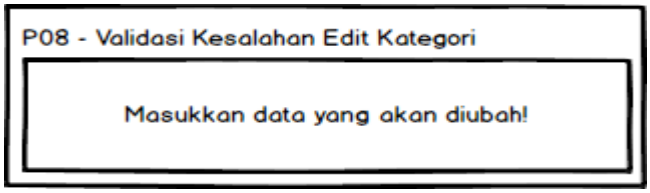
Gambar III.53 Perancangan Pesan P05 – Validasi Kesalahan Tambah Situs



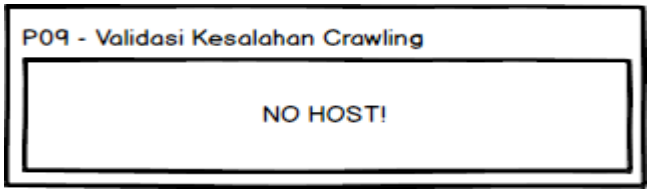
Gambar III.54 Perancangan Pesan P06 – Validasi Kesalahan Edit Situs



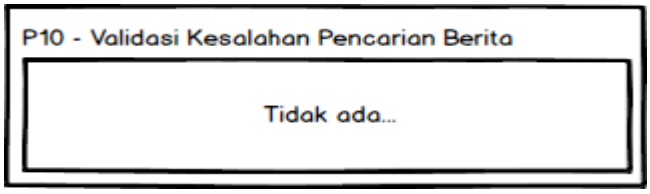
Gambar III.55 Perancangan Pesan P07 – Validasi Kesalahan Tambah Kategori



Gambar III.56 Perancangan Pesan P08 – Validasi Kesalahan Edit Kategori



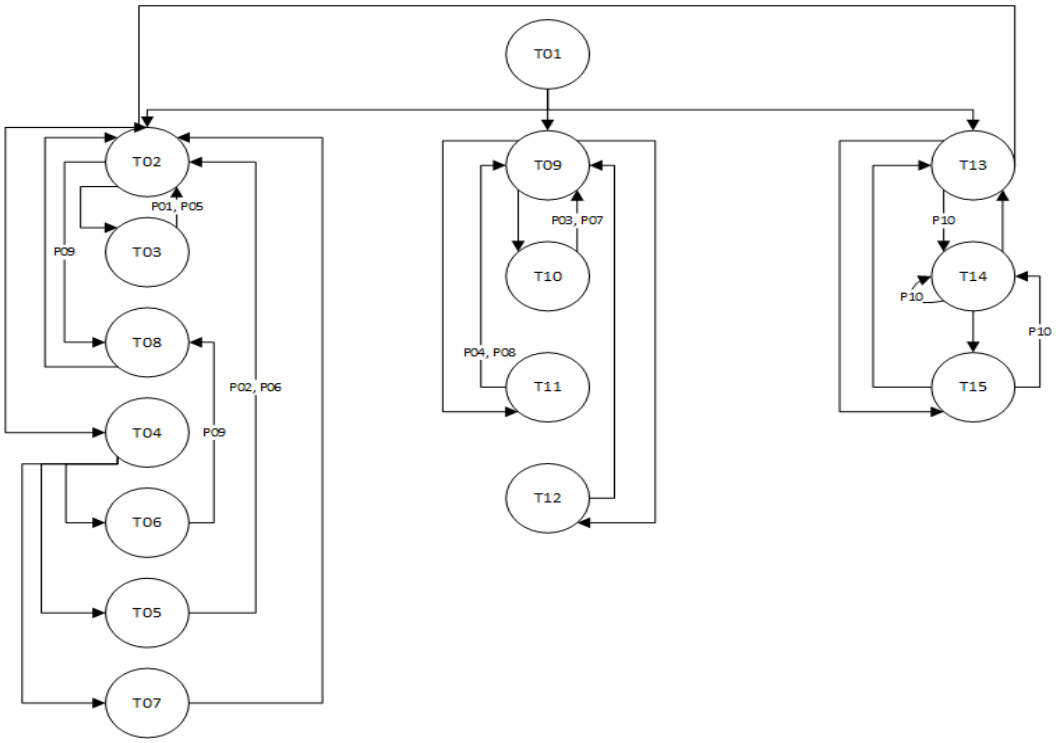
Gambar III.57 Perancangan Pesan P09 – Validasi Kesalahan *Crawling*



Gambar III.58 Perancangan Pesan P10 – Validasi Kesalahan Pencarian Berita

III.10 Jaringan Semantik

Jaringan semantik menggambarkan keterhubungan dari setiap menu dari tampilan satu ke tampilan lain. Berikut adalah jaringan semantik pada sistem *news aggregator* :



Gambar III.59 Jaringan Semantik