

BAB III

METODE PENELITIAN DAN PERANCANGAN SISTEM

3.1 Analisa Permasalahan Sistem

Berdasarkan pada latar belakang yang dihadapi, dibutuhkan suatu aplikasi yang dapat digunakan untuk mengambil berita online yang terdapat pada suatu situs berita serta mengklasifikasikannya secara otomatis. Pengembangan aplikasi internet agent yang dapat mengklasifikasikan berita sebagai solusi, dapat menghemat waktu pencarian berita di internet. Pengguna tidak perlu langsung mengakses situs penyedia berita yang umumnya dipenuhi oleh banner sehingga membuat akses untuk membaca berita menjadi cukup lambat, apalagi jika bandwidth yang digunakan terbatas.

Setiap provider situs umumnya memiliki struktur situs yang berbeda-beda. Maka dari itu diperlukan waktu yang cukup untuk menganalisa struktur situs tersebut. Untuk menganalisa struktur situs tersebut, diperlukan halaman berita dari provider tersebut, kemudian dipelajari tag-tag apa yang mengapit data-data berita yang dibutuhkan seperti judul, link, waktu dan isi berita tersebut. Setelah itu baru dibuat pola regex yang dapat mengambil data-data berita tersebut. Kesulitan yang dialami adalah apabila struktur dari situs tersebut setiap saat berubah.

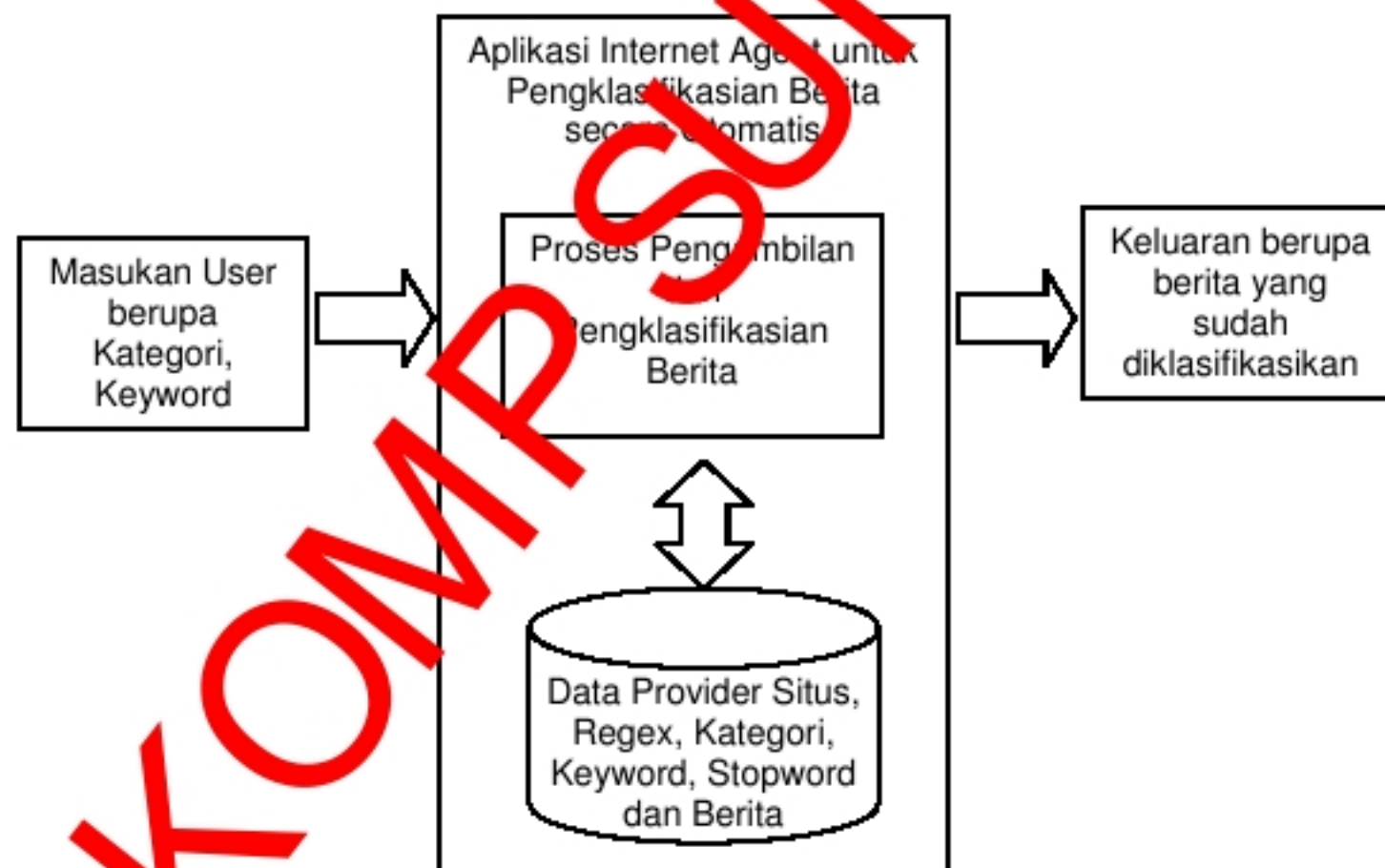
Proses pengambilan berita dilakukan dengan cara membaca dokumen HTML situs berita tersebut dan mengekstrak informasi yang dibutuhkan, seperti judul berita, link, waktu publikasi berita serta isi berita, dengan menggunakan pola regex yang telah dibuat sesuai dengan struktur dari situs tersebut. Setelah berita di dapat

kemudian dilanjutkan dengan proses pengklasifikasian. Metode yang digunakan dalam proses pengklasifikasian berita adalah Fuzzy Search dan Association Rule.

3.2 Perancangan Sistem

3.2.1 Desain Umum Sistem

Desain umum aplikasi yang dibuat digambarkan dengan Block Diagram. Block Diagram menggambarkan alur sistem secara umum yang berupa masukan, proses dan keluaran dari sistem. Desain umum dari aplikasi ini dapat dilihat pada gambar 3.1.



Gambar 3.1 Blok Diagram Aplikasi Internet Agent

Pada gambar diatas dijelaskan proses pengambilan dan pengklasifikasian berita yang terjadi di dalam sistem. Pertama, sistem menerima inputan berupa

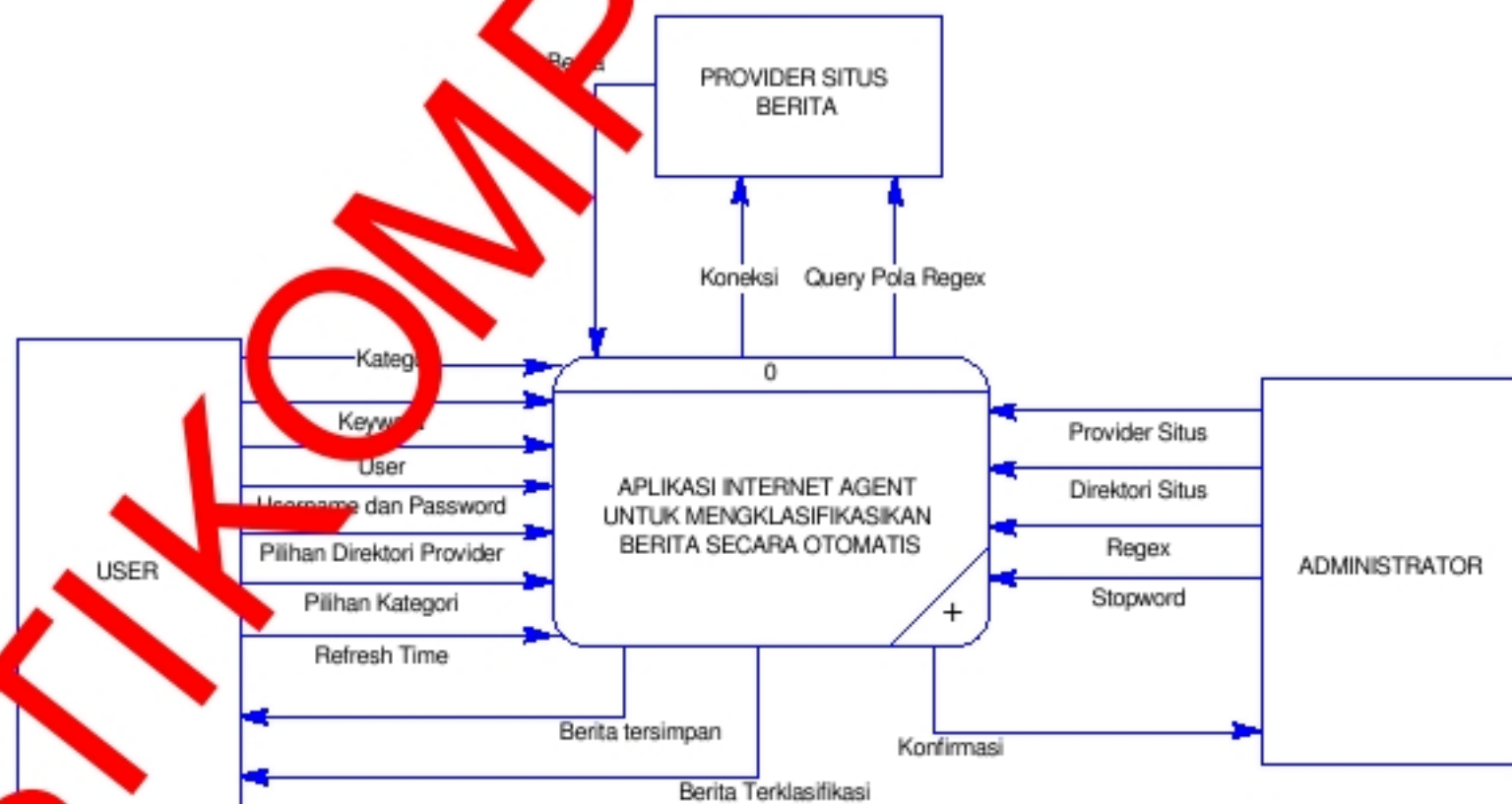
kategori dan keyword kemudian akan dilakukan proses pencarian keyword pada isi berita menggunakan Metode Fuzzy Search. Pengklasifikasiannya sendiri menggunakan Metode Association Rule. Keluaran dari aplikasi ini berupa berita-berita yang sudah diklasifikasikan sesuai dengan kategori yang diinputkan.

3.2.2 DFD

Pengembangan dari sistem ini akan menggunakan Data Flow Diagram sebagai media untuk menjelaskan semua alur data beserta proses-proses yang terdapat di dalam sistem. Berikut ini *context diagram* dari aplikasi internet agent untuk mengklasifikasikan berita secara otomatis, DFD Level 0 dan sub-sub prosesnya.

A. Context Diagram

Context diagram dari aplikasi ini dapat dilihat pada gambar 3.2.



Gambar 3.2 Context Diagram

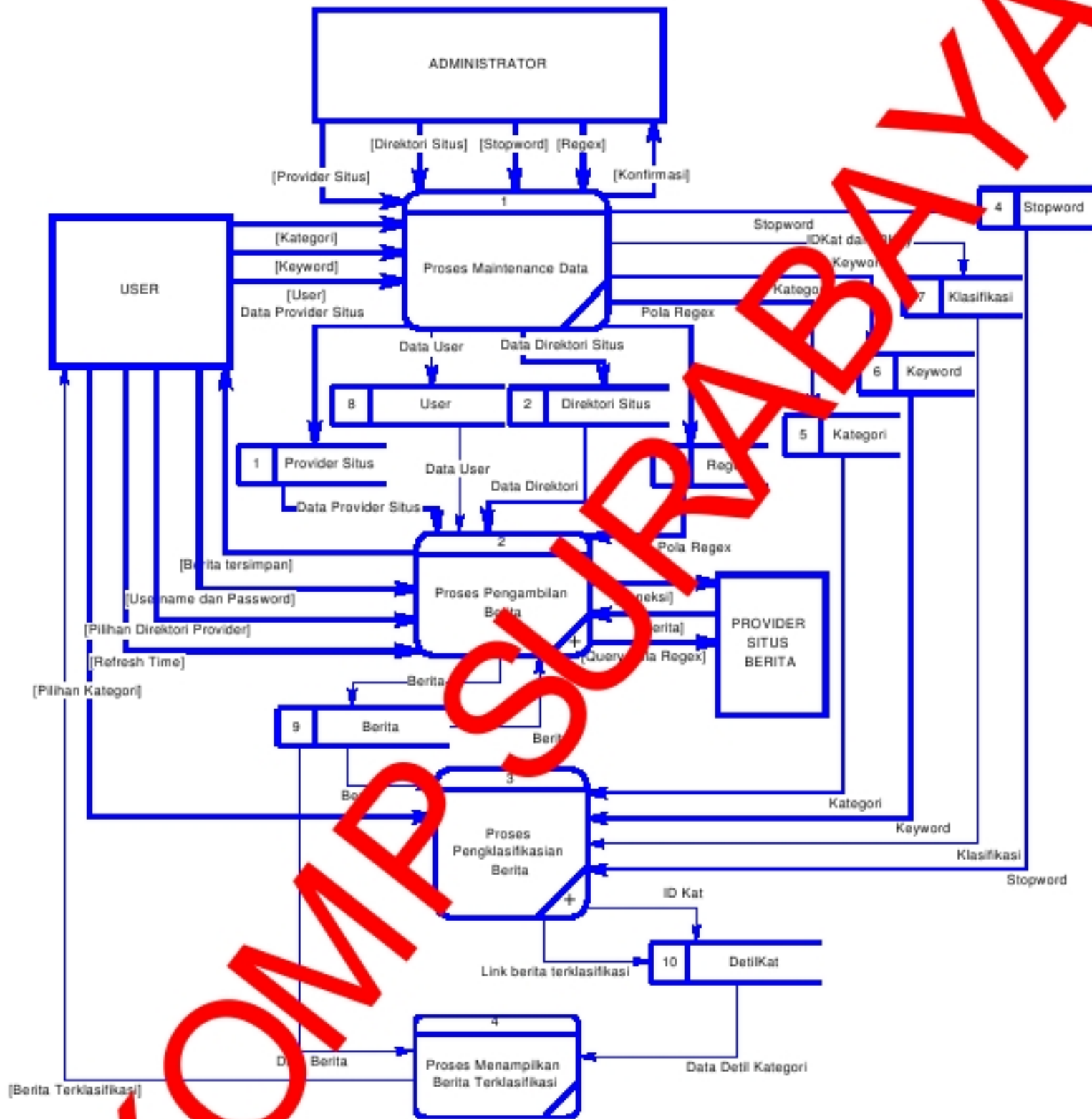
Context Diagram Aplikasi Internet Agent pada gambar 3.2, memiliki 3 entitas luar yaitu User, Administrator dan Provider Situs Berita. Penjelasan dari gambar diatas akan dijelaskan dibawah ini.

Penjelasan :

1. Entity User memberikan 7 inputan yaitu data kategori, keyword, user, username dan password, pilihan kategori, pilihan direktori situs dan refresh time
2. Entity Administrator memberikan inputan berupa data provider situs, direktori situs, regex dan stopword. Inputan ini dilakukan saat pertama kali pembuatan program.
3. Entity Provider Situs menerima inputan koneksi, query pola regex dan memberikan inputan ke proses berita.

Setelah *context diagram* dekomposisi akan didapat DFD level 0 seperti yang terlihat pada gambar 3.3 yang terdiri dari 4 proses.

B. DFD Level 0



Gambar 3.3 DFD Level 0

DFD level 0 pada gambar 3.3 diatas terdapat 4 proses dan 9 database. Ke

empat proses itu yaitu :

1. Proses *Maintenance* Data. Proses ini digunakan untuk *me-maintenance* data *user*, kategori, keyword, provider situs, direktori situs. Proses ini akan menghasilkan keluaran berupa data konfirmasi ke administrator.
2. Proses Pengambilan Berita. Proses ini digunakan untuk melakukan proses pengambilan berita ke provider situs yang sudah ditentukan. Proses ini menerima inputan dari user berupa username dan password, refresh time dan pilihan direktori provider. Proses ini dapat didekomposisi lagi seperti terlihat pada Gambar 3.4.
3. Proses Pengklasifikasian Berita. Proses ini digunakan untuk melakukan proses pengklasifikasian berita. Proses ini menerima inputan dari Proses Pengambilan Berita berupa array berita. Selain itu proses ini juga menerima inputan dari user berupa pilihan kategori. User akan menerima output dari proses ini berupa berita yang sudah terklasifikasi. Proses ini dapat didekomposisi lagi seperti terlihat pada Gambar 3.5.
4. Proses Menampilkan Berita Terklasifikasi. Proses ini digunakan untuk menampilkan berita-berita yang sudah terklasifikasi. Proses ini menerima inputan dari database detail kategori dan berita.

Sembilan database itu adalah sebagai berikut :

1. User, digunakan untuk menyimpan data user.
2. Provider Situs, digunakan untuk menyimpan data provider situs.
3. Direktori Provider, digunakan untuk menyimpan data direktori provider.
4. Regex, digunakan untuk menyimpan data pola regex.
5. Stopword, digunakan untuk menyimpan stopword.

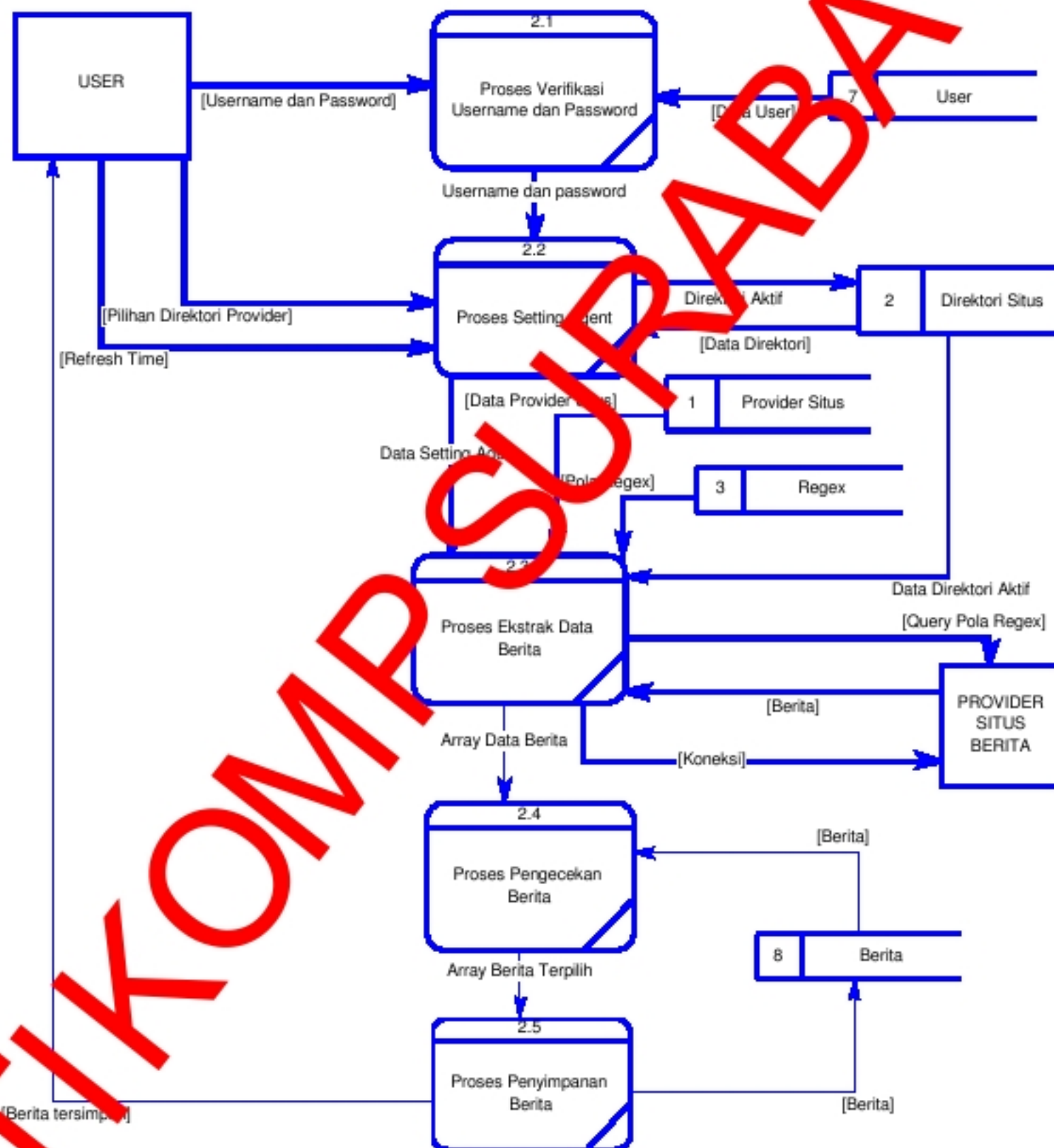
6. Kategori, digunakan untuk menyimpan data kategori.
7. Keyword, digunakan untuk menyimpan keyword.
8. Berita, digunakan untuk menyimpan berita yang didapat dari provider situs.
9. Detil Kategori, digunakan untuk menyimpan hasil klasifikasi.

C. DFD Level 1 Sub Proses Pengambilan Berita

Pada DFD Level 1 Sub Proses Pengambilan berita terdapat 4 proses yaitu :

1. Proses Verifikasi Username dan Password. Proses ini harus dilakukan terlebih dahulu jika ingin masuk ke proses setting agent. Hanya user yang sudah terdaftar saja yang boleh melakukan proses internet agent dan klasifikasi.
2. Proses Setting Agent. Proses ini digunakan untuk melakukan setting awal untuk proses pengambilan berita. Di sini user memasukkan refresh time dan direktori provider situs yang dipilih.
3. Proses Ekstrak Data Berita. Proses ini digunakan untuk melakukan proses pengambilan informasi berita berupa link, judul, waktu dan isi berita. Proses ini menerima inputan dari Proses Setting Agent berupa data setting agent. Proses ini juga menerima inputan dari database provider situs, direktori situs dan regex. Proses ini akan menerima data-data berita dari Provider Situs.
4. Proses Pengecekan Berita. Proses ini digunakan untuk melakukan pengecekan apakah berita yang diterima sudah ada pada database atau tidak. Proses ini akan mengeluarkan output berupa data-data berita yang belum ada pada database.

5. Proses Penyimpanan Berita. Proses ini digunakan untuk melakukan penyimpanan berita ke database berita. Inputannya berupa array data berita yang sudah melalui Proses Pengecekan Berita terlebih dahulu.

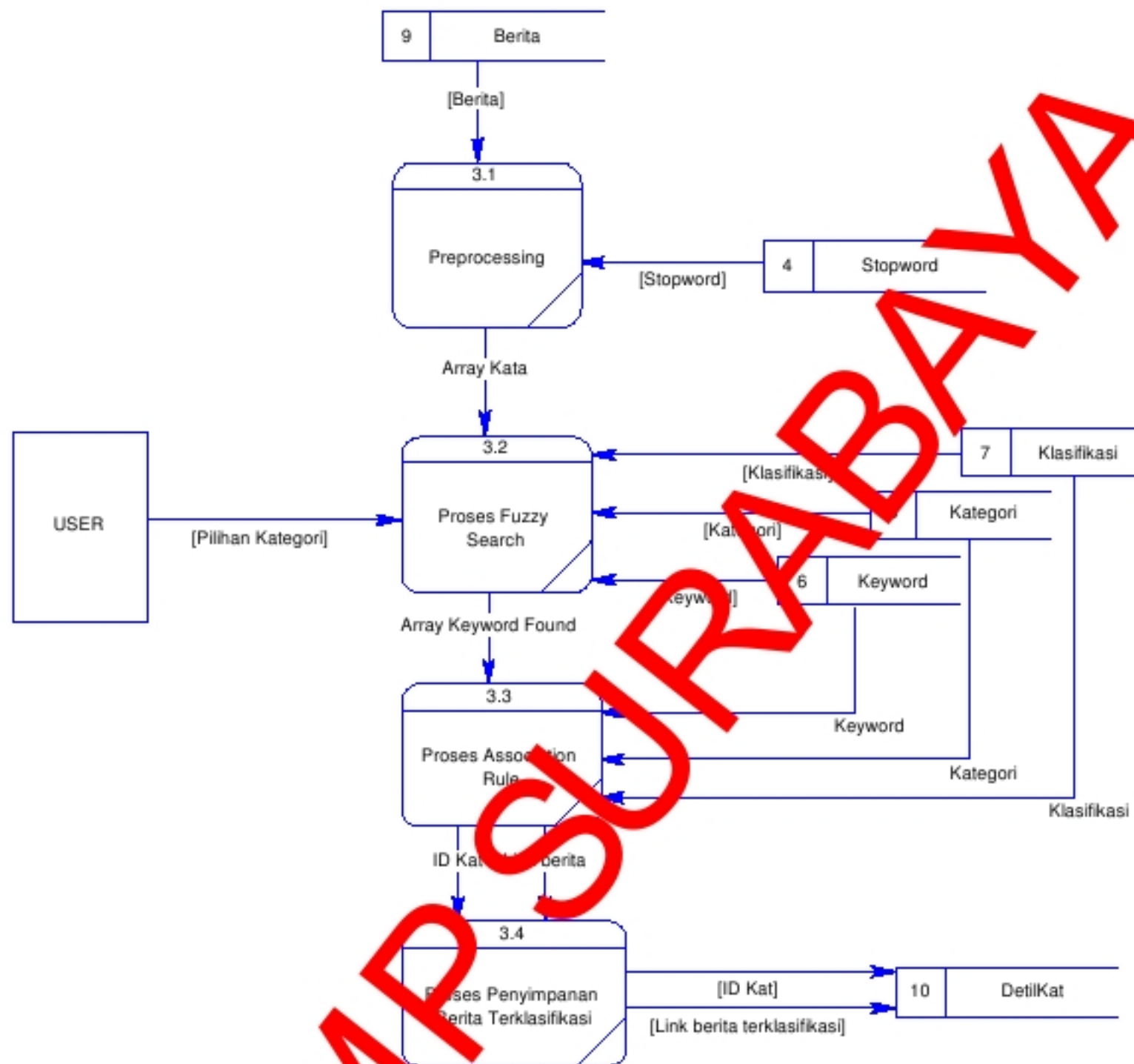


Gambar 3.4 DFD Level 1 Sub Proses Pengambilan Berita

D. DFD Level 1 Sub Proses Pengklasifikasian Berita

Pada DFD Level 1 Sub Proses Pengklasifikasian Berita yang dapat dilihat pada Gambar 3.5, terdapat 4 proses yaitu :

1. *Preprocessing*. Proses ini digunakan untuk melakukan tahap awal pada proses pengklasifikasian. Proses ini menerima inputan dari database stop word berupa kata-kata yang diabaikan dalam proses pengklasifikasian serta kata sambung. Dalam Proses ini akan dilakukan penghapusan tanda baca, kata sambung serta pemisahan kata.
2. Proses Fuzzy Search. Proses ini digunakan untuk melakukan proses *search and matching* kata-kata pada isi berita dengan keyword-keyword yang sesuai dengan kategori yang dipilih menggunakan metode Fuzzy Search. User memberikan inputan berupa pilihan kategori. Proses ini juga menerima inputan dari database kategori dan keyword.
3. Proses Association Rule. Proses ini digunakan untuk melakukan pengklasifikasian dengan metode Association Rule. Proses ini menerima inputan dari Proses Fuzzy Search berupa keyword-keyword yang ditemukan pada isi berita. Proses ini juga menerima inputan dari database kategori dan keyword. Output dari proses ini adalah berita yang sudah diklasifikasikan.
4. Proses Penyimpanan Berita Terklasifikasi. Proses ini digunakan untuk melakukan penyimpanan berita sesuai dengan kategorinya pada database Detil Kategori.



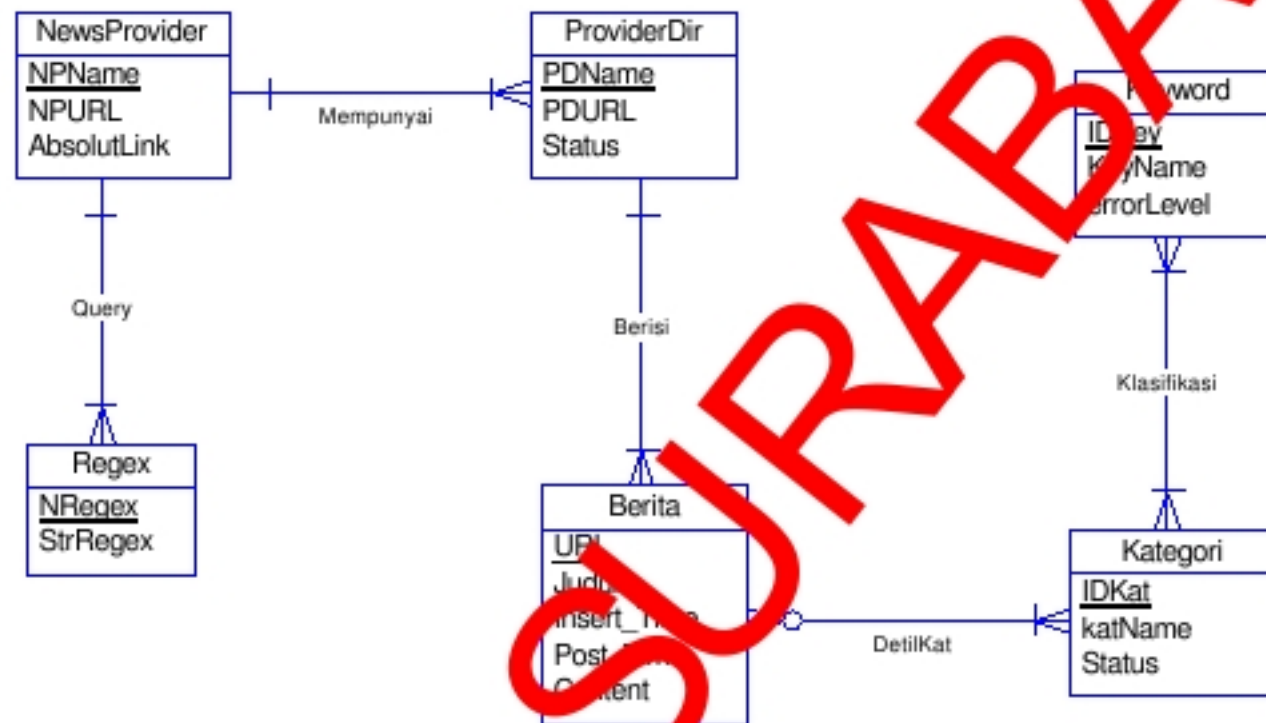
Gambar 3.2 DFD Level 1 Sub Proses Pengklasifikasian Berita

3.2.3 ERD

ERD menunjukkan struktur keseluruhan kebutuhan data serta hubungan data-data yang digunakan dalam sistem. Dalam ERD, data tersebut digambarkan dengan menggunakan simbol entitas. Dalam perancangan sistem ini terdapat beberapa entitas yang saling terkait. Untuk membuat ERD digunakan perangkat lunak *PowerDesigner DataArchitect* yang disajikan dalam bentuk *Conceptual Data Model*

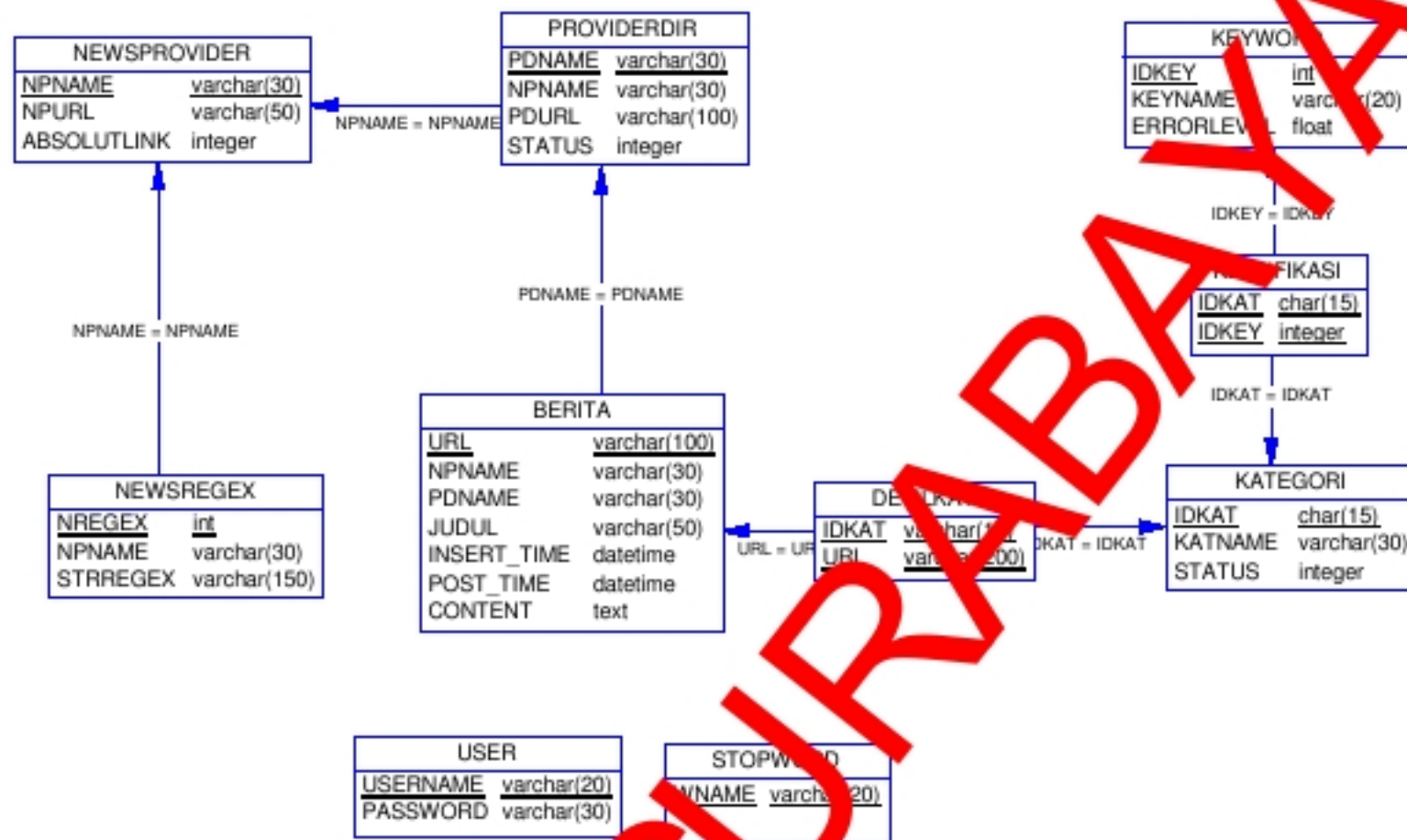
(CDM) dan Physical Data Model (PDM). Berikut ini ERD dari aplikasi internet agent yang dibuat.

A. CDM



Gambar 3.6 ERD CDM

B. PDM



Gambar 3.7 ERD PDM

3.2.4 Struktur database

Rincian dari ERD digambarkan dengan struktur *database* yang terdiri dari kolom-kolom yang memiliki atribut berupa nama kolom, tipe data, batasan atau aturan yang mengarah pada tabel tertentu dan keterangan. Struktur *database* menunjukkan daftar kebutuhan tabel yang digunakan untuk menyimpan data yang diperlukan dalam sistem ini.

Berikut tabel-tabel yang terlibat dalam aplikasi internet agent yang dibuat :

1. Tabel : **NewsProvider**

Fungsi : Menyimpan data provider situs berita (tabel 3.1)

Tabel 3.1 Tabel News Provider

Nama Kolom	Tipe Data	Batasan	Keterangan
NPName	Varchar(30)	PK	Nama Provider
NPURL	Varchar(50)		URL Provider
AbsolutLink	Integer		AbsolutLink

2. Tabel : NewsDirektori

Fungsi : Menyimpan data direktori provider situs (tabel 3.2)

Tabel 3.2 Tabel NewsDirektori

Nama Kolom	Tipe Data	Batasan	Keterangan
PDName	Varchar(30)	PK	Nama Direktori
NPName	Varchar(30)	FK	Nama Provider
PDURL	Varchar(100)		URL Direktori
Status	Integer		Status

3. Tabel : Regex

Fungsi : Menyimpan pola regex (tabel 3.3)

Tabel 3.3 Tabel Regex

Nama Kolom	Tipe Data	Batasan	Keterangan
NRegex	Integer	PK	Tipe Regex
NPName	Varchar(30)	PK	Nama Provider
StRegex	Varchar(150)		Pola Regex

4. Tabel : Berita

Fungsi : Menyimpan berita (tabel 3.4)

Tabel 3.4 Tabel Berita

Nama Kolom	Type Data	Batasan	Keterangan
URL	Varchar(200)	PK	URL Berita
NPName	Varchar(30)	FK	Nama Provider
PDName	Varchar(30)	FK	Nama Direktori
Judul	Varchar(100)		Judul berita
Insert_Time	DateTime		Waktu masuk
Post_Time	DateTime		Waktu berita
Content	Text		Isi berita

5. Tabel : Kategori

Fungsi : Menyimpan Kategori (tabel 3.5)

Tabel 3.5 Tabel Kategori

Nama Kolom	Type Data	Batasan	Keterangan
IDKat	Varchar(10)	PK	Kode Kategori
KatName	Varchar(20)		Nama Kategori
Status	Integer		Status

6. Tabel : Keyword

Fungsi : Menyimpan keyword (tabel 3.6)

Tabel 3.6 Tabel Keyword

Nama Kolom	Type Data	Batasan	Keterangan
IDKey	Integer	PK	Kode Keyword
KeyName	Varchar(20)		Keyword
ErrorLevel	Float		ErrorLevel

7. Tabel : Klasifikasi

Fungsi : Menyimpan data klasifikasi (tabel 3.7)

Tabel 3.7 Tabel Klasifikasi

Nama Kolom	Tipe Data	Batasan	Keterangan
IDKat	Varchar(10)	PK	Kode Kategori
IDKey	Integer	PK	Kode Keyword

8. Tabel : DetilKat

Fungsi : Menyimpan data detil kategori (tabel 3.8)

Tabel 3.8 Tabel DetilKat

Nama Kolom	Tipe Data	Batasan	Keterangan
IDKat	Varchar(10)	PK	Kode Kategori
URL	Varchar(200)	PK	URL Berita

9. Tabel : User

Fungsi : Menyimpan data user (tabel 3.9)

Tabel 3.9 Tabel User

Nama Kolom	Tipe Data	Batasan	Keterangan
Username	Varchar(20)	PK	Username
Password	Varchar(30)		Password

10. Tabel : Stop Word

Fungsi : Menyimpan kata stopword (tabel 3.10)

Tabel 3.10 Tabel StopWord

Nama Kolom	Tipe Data	Batasan	Keterangan
WName	Varchar(15)	PK	Kata stopword

3.2.5 Desain user interface

A. Halaman Administrator

Halaman ini digunakan untuk melakukan *maintenance* data yang digunakan dalam aplikasi ini. Seperti Provider Situs, Direktori Situs, Regex dan Stopword. Yang bisa mengakses halaman ini hanya administrator. Berikut digambarkan rancangan user interface dari form-form yang digunakan pada halaman ini.

1. Form Login

Form ini digunakan untuk login sebagai administrator.

Form Login	
Username	Varchar(20)
Password	Varchar(30)
Login	

Gambar 3.8 Form Login Admin

2. Halaman Menu Admin

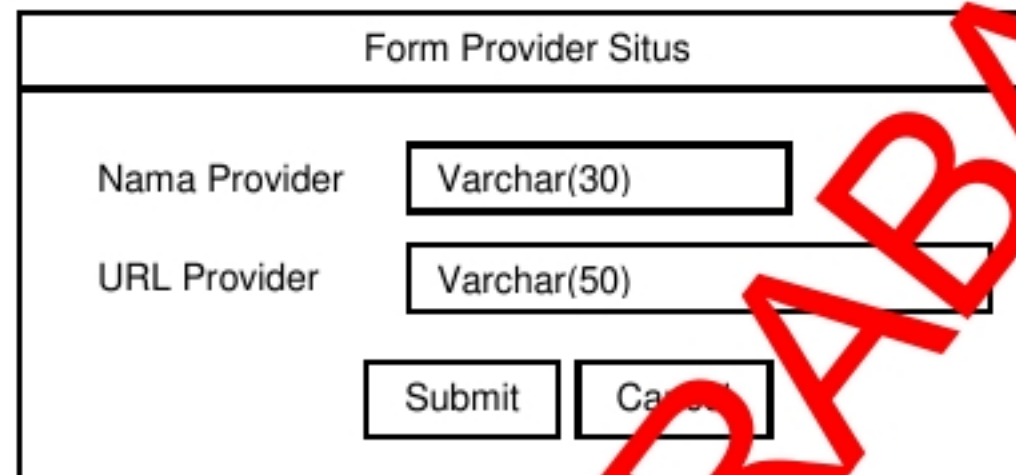
Halaman ini akan menampilkan menu-menu yang bisa digunakan oleh administrator.

Menu Administrator		
[Menu 1]	[Menu 2]	[Menu 3]
[Menu 4]	[Menu 5]	[Menu 6]

Gambar 3.9 Halaman Menu Administrator

3. Form Provider Situs

Form ini digunakan untuk memasukkan data provider situs yang sudah ditentukan sebelumnya.

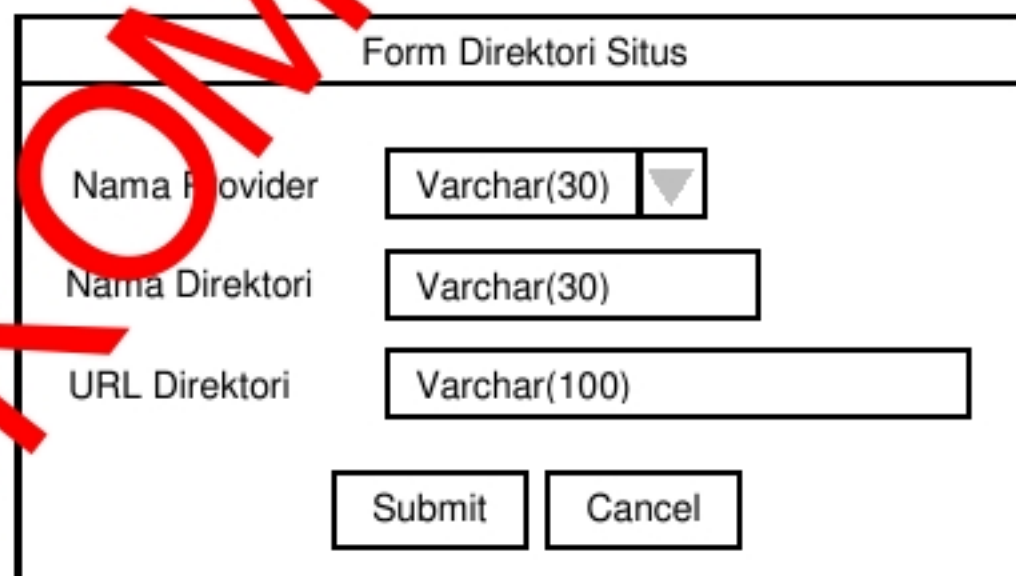


Form Provider Situs	
Nama Provider	<input type="text" value="Varchar(30)"/>
URL Provider	<input type="text" value="Varchar(50)"/>
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.10 Form Provider Situs

4. Form Direktori Situs

Form ini digunakan untuk memasukkan data direktori dari provider situs yang telah dimasukkan sebelumnya.

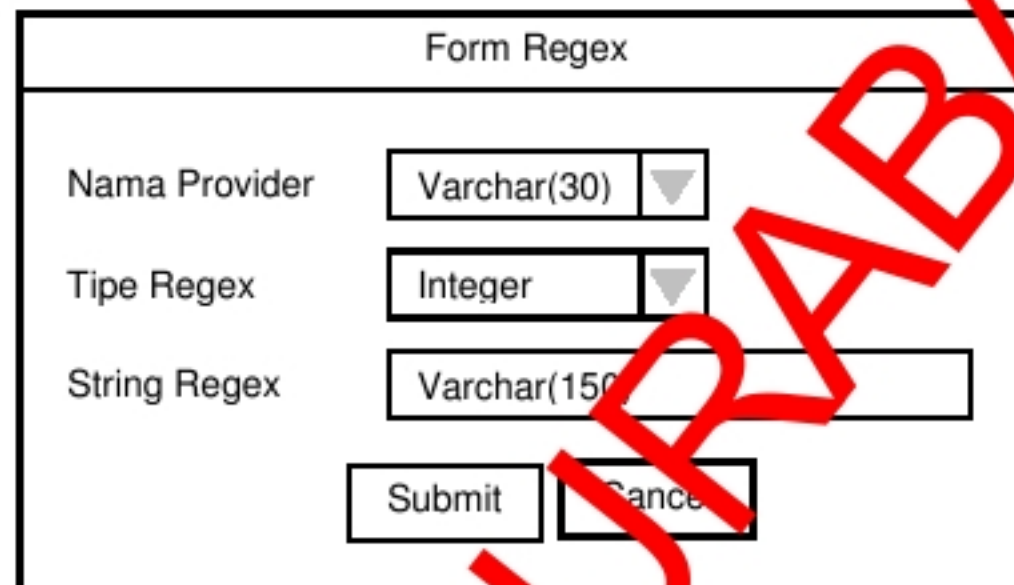


Form Direktori Situs	
Nama Provider	<input type="text" value="Varchar(30)"/> ▼
Nama Direktori	<input type="text" value="Varchar(30)"/>
URL Direktori	<input type="text" value="Varchar(100)"/>
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.11 Form Direktori Situs

5. Form Regex

Form ini digunakan untuk memasukkan pola regex yang telah dibuat. Dimana tipe regex berisi nilai 1-3 yang mendefinisikan 3 type regex yang digunakan yaitu posisi, data berita dan isi berita.

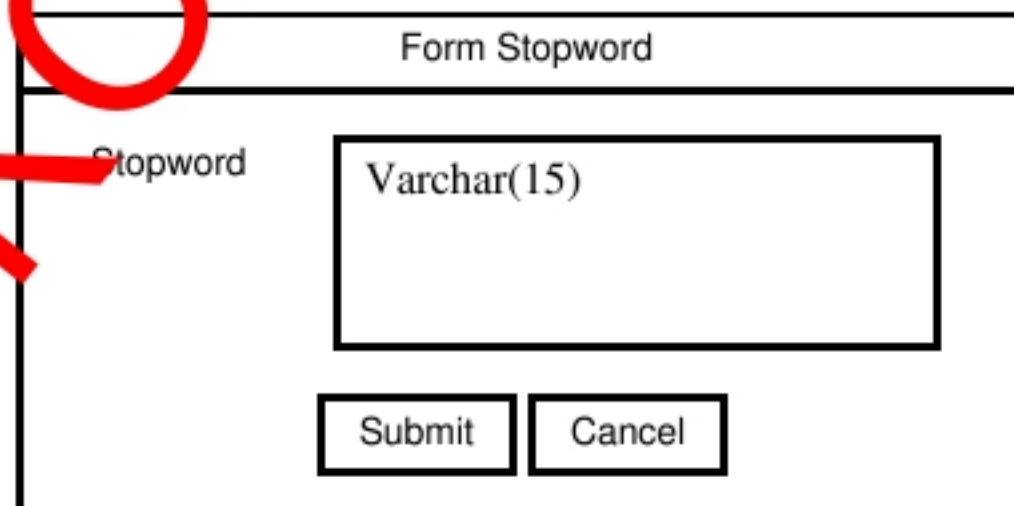


Form Regex	
Nama Provider	Varchar(30) ▼
Tipe Regex	Integer ▼
String Regex	Varchar(150)
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.12 Form Regex

6. Form Stopword

Form ini digunakan untuk memasukkan stopwords atau kata-kata yang diabaikan dalam proses penklasifikasian seperti kata sambung.



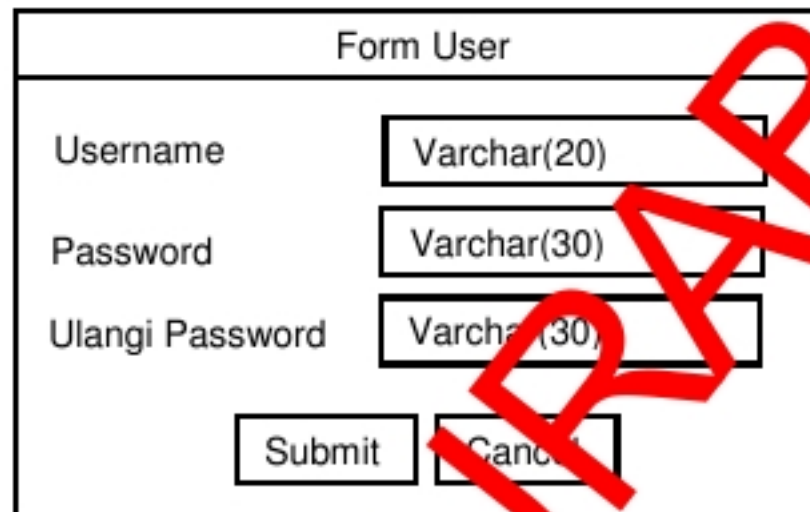
Form Stopword	
Stopword	Varchar(15)
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.13 Form Stopword

B. Halaman User

1. Form User

Form ini digunakan untuk memasukkan data user yang berupa username dan password.

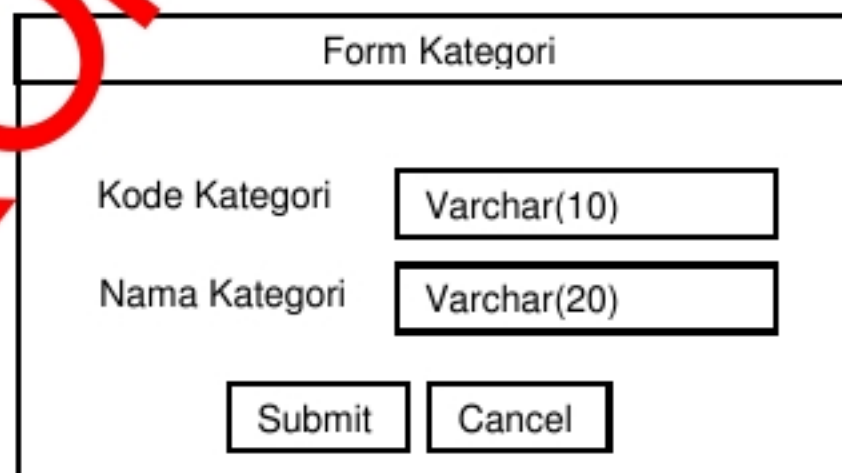


Form User	
Username	Varchar(20)
Password	Varchar(30)
Ulangi Password	Varchar(30)
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.14 Form User

2. Form Kategori

Form ini digunakan untuk media untuk menambah kategori baru dimana kode kategori di *generate* secara otomatis.



Form Kategori	
Kode Kategori	Varchar(10)
Nama Kategori	Varchar(20)
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.15 Form Kategori

3. Form Keyword

Form ini digunakan untuk media untuk menambah keyword.

Form Keyword	
Kode Kategori	KXXXX
Nama Kategori	XXXXXX
Keyword	Varchar(20)
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	

Gambar 3.16 Form Kategori

4. Form Login

Form ini digunakan untuk login sebagai user

Form Login	
Username	Varchar(20)
Password	Varchar(30)
<input type="button" value="Login"/>	

Gambar 3.17 Form Kategori

5. Form Ganti Password

Form ini digunakan untuk melakukan penggantian password.

Form Ganti Password

Username

Password

Password Baru

Ulangi Password

Gambar 3.18 Form Ganti Password

6. Halaman Setting Agent

Form ini digunakan untuk memilih direktori provider sebagai tempat pencarian berita yang diinginkan. Jika fungsi Aktif berarti akan dilakukan pencarian pada direktori tersebut. Dan jika Pasif berarti sebaliknya.

[Menu]

Provider	Direktori	URL	Edit	Fungsi
Provider	Direktori	URL	[Edit]	[Aktif]
Provider	Direktori	URL	[Edit]	[Pasif]

Gambar 3.19 Halaman Setting Agent

7. Halaman Setting Kategori

Form ini digunakan untuk memilih kategori yang ingin di cari.

[Menu]

ID Kat	Kategori	Detil	Fungsi
ID Kat	Kategori	[Detil]	[Aktif]
ID Kat	Kategori	[Detil]	[Aktif]

<< BACK NEXT >>

Gambar 3.20 Halaman Setting Kategori

8. Halaman Detil Kategori

Halaman ini digunakan untuk menampilkan data detil kategori.

[Menu]

Kode Kategori : KXXXX

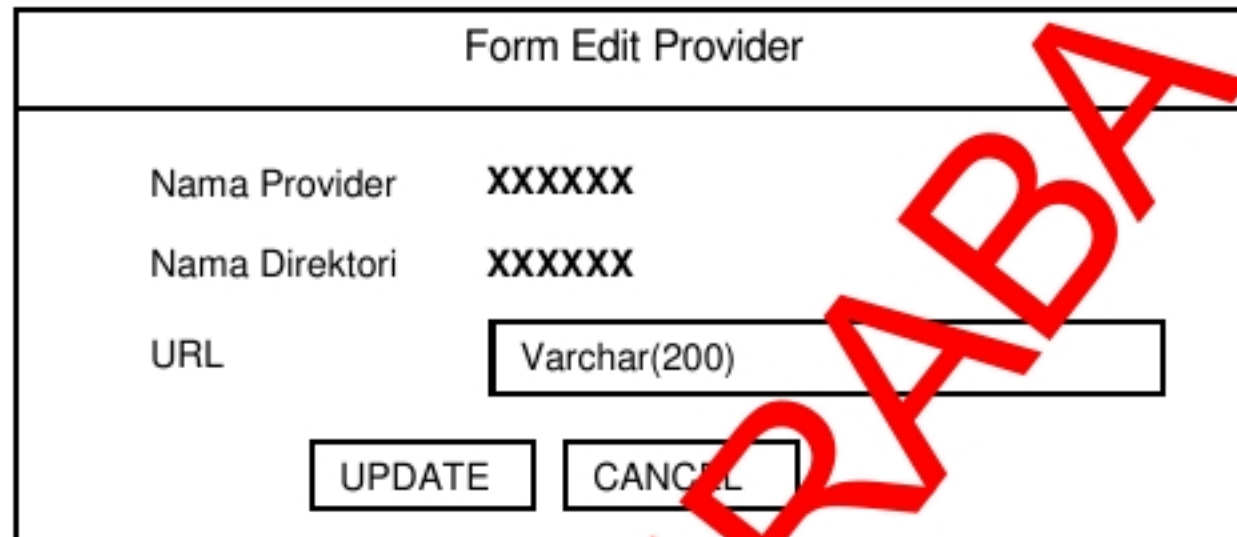
Nama Kategori : XXXXX

Keyword	Error Level	Edit	Delete
ID Kat	Nilai Error	[Edit]	[Delete]
ID Kat	Nilai Error	[Edit]	[Delete]

Gambar 3.21 Halaman Detil Kategori

9. Form Edit Direktori Provider

Form ini digunakan untuk mengedit url direktori provider jika terjadi perubahan alamat.

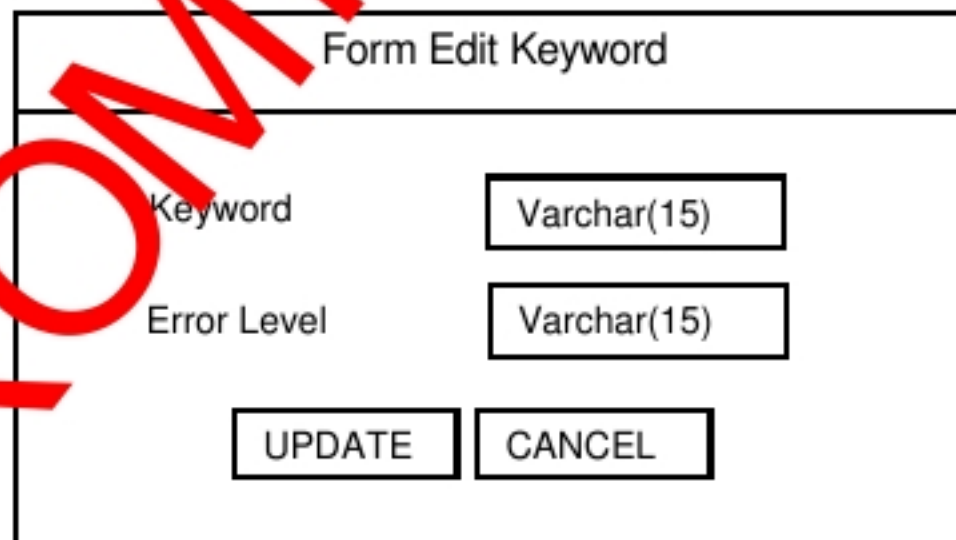


Form Edit Provider	
Nama Provider	XXXXXX
Nama Direktori	XXXXXX
URL	<input type="text" value="Varchar(200)"/>
<input type="button" value="UPDATE"/> <input type="button" value="CANCEL"/>	

Gambar 3.22 Form Edit Direktori Provider

10. Form Edit Keyword

Form ini digunakan untuk mengedit keyword dan error levelnya.

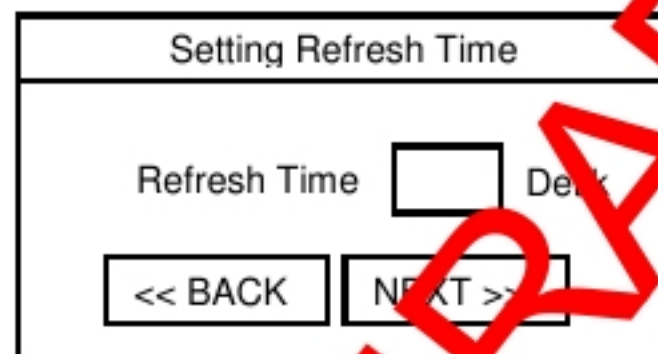


Form Edit Keyword	
Keyword	<input type="text" value="Varchar(15)"/>
Error Level	<input type="text" value="Varchar(15)"/>
<input type="button" value="UPDATE"/> <input type="button" value="CANCEL"/>	

Gambar 3.23 Form Edit Keyword

11. Halaman Setting Refresh Time

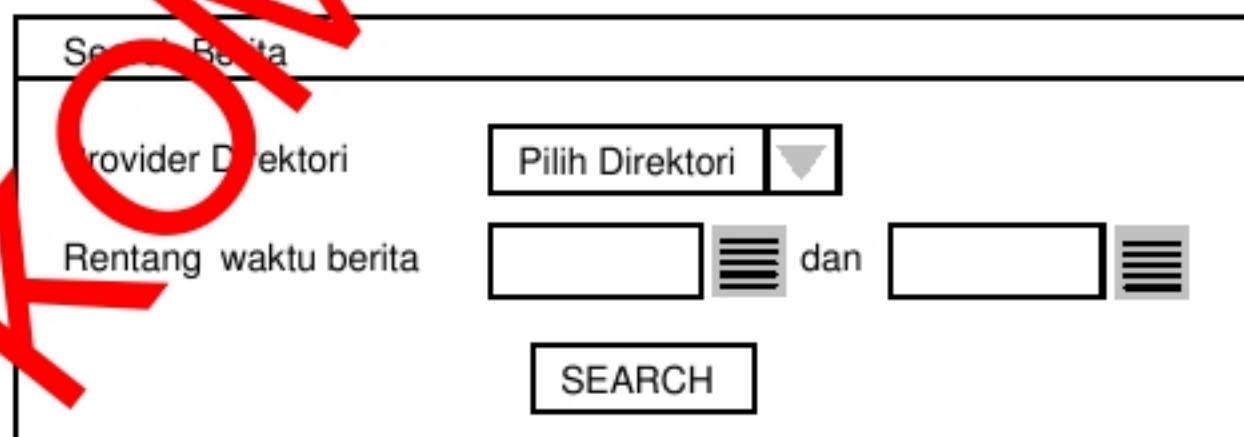
Halaman ini digunakan untuk melakukan setting refresh time. Refresh time disini adalah waktu yang ditetapkan user, tiap berapa detik user ingin aplikasi internet agent ini melakukan pencarian berita.



Gambar 3.24 Halaman Setting Refresh Time

12. Halaman Search Berita

Halaman ini digunakan untuk melakukan pencarian berita berdasarkan direktori provider dan rentang waktu berita.



Gambar 3.25 Halaman Search Berita

13. Halaman Maintenance Berita

Halaman ini digunakan untuk menghapus berita.

[First] [Prev]	1 2 3 4 5 6 ...				[Next] [Last]
No	Sumber	Judul	Waktu	Delete	
1	Sumber	Judul	Waktu	[Delete]	
2	Sumber	Judul	Waktu	[Delete]	
[First] [Prev]	1 2 3 4 5 6 ...				[Next] [Last]

Gambar 3.26 Halaman Maintenance Berita

3.3 Desain Uji Coba

Uji coba validasi sistem bertujuan untuk memastikan bahwa sistem telah dibuat dengan benar sesuai dengan kebutuhan atau tujuan yang diharapkan. Kekurangan atau kelemahan sistem pada tahap ini akan dievaluasi sebelum diimplementasikan secara riil.

Proses pengujian akan dilakukan dengan melakukan berbagai percobaan untuk membuktikan bahwa aplikasi yang telah dibuat sesuai dengan tujuan, antara lain (1) uji implementasi pengambilan berita pada situs berita, (2) uji implementasi *fuzzy search*, (3) uji implementasi pengklasifikasian dengan *association rule*, (4) uji aplikasi internet agent secara keseluruhan. Untuk uji coba aplikasi secara keseluruhan akan diuji menggunakan Black Box Texting.

3.3.1 Uji implementasi pengambilan berita

Proses uji coba ini dilakukan dengan membandingkan jumlah berita yang ada pada halaman berita dari provider situs dengan jumlah berita yang dapat diambil oleh aplikasi internet agent. Uji coba ini dapat dikatakan berhasil jika jumlah dan isi berita yang dapat diambil sama dengan jumlah dan isi berita pada provider situs berita.

3.3.2 Uji implementasi fuzzy search

Proses uji coba ini dilakukan dengan cara membandingkan kata-kata yang ada pada isi berita suatu berita dengan keyword-keyword dari kategori yang dipilih. Apabila pada implementasinya kata-kata yang ditemukan sesuai dengan keyword-keyword ataupun dapat menemukan variasi variasi kata dari keyword tersebut, maka uji coba ini dapat dikatakan berhasil.

3.3.3 Uji implementasi pengklasifikasian dengan association rule

Tujuan dari uji coba ini adalah untuk mengetahui apakah suatu berita sudah dapat diklasifikasikan dengan benar sesuai dengan yang diharapkan. Untuk uji coba ini akan dipilih satu berita untuk diklasifikasikan. Kemudian akan diklasifikasikan dengan menggunakan metode Association Rule. Jika berita tersebut dapat diklasifikasikan sesuai dengan kategori yang diharapkan, maka uji coba ini dapat dikatakan berhasil.

3.3.4 Uji aplikasi internet agent secara keseluruhan

Proses uji coba ini dilakukan untuk mengetahui apakah fungsi klasifikasi secara keseluruhan sudah berjalan sesuai dengan tujuan yang diharapkan. Untuk uji coba ini akan dipilih beberapa berita untuk diklasifikasikan. Untuk menguji kualitas dari solusi yang dihasilkan, akan dianalisa berdasarkan perbandingan antara jumlah solusi yang seharusnya benar dengan total jumlah permasalahan yang dihadapkan (Quast, 2001). Maka dari itu untuk mendapatkan akurasi atau kualitas dari klasifikasi berita yang dihasilkan digunakan rumus sebagai berikut

$$\text{Akurasi} = \frac{\text{Total klasifikasi berita yang benar}}{\text{Total jumlah berita}} \times 100\% \dots\dots\dots(5)$$

STIKOMPSURABAYA