



**IMPLEMENTASI EKSTRAKSI WEB (*WEB SCRAPING*) PADA SITUS
BERITA MENGGUNAKAN METODE EKSPRESI REGULER**

SKRIPSI

Oleh

**Muhamat Abdul Rohim
NIM 1424101010**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER**

2018



**IMPLEMENTASI EKSTRAKSI WEB (*WEB SCRAPING*) PADA SITUS
BERITA MENGGUNAKAN METODE EKSPRESI REGULER**

SKRIPSI

Diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk menyelesaikan Pendidikan Sarjana (S1) Program Studi Sistem Informasi Universitas Jember dan mencapai gelar Sarjana Komputer

Oleh

Muhamat Abdul Rohim
NIM 1424101010

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER

2018

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya untuk mempermudah dan melancarkan dalam pengerjaan skripsi;
2. Ayahanda Muki, ibunda Rohmani, dan nenek saya Alm. Muntini;
3. Sahabatku bersama dukungan dan doanya;
4. Guru - guruku sejak sekolah dasar sampai dengan perguruan tinggi;
5. Almamater Program Studi Sistem Informasi Universitas Jember dan Fakultas Ilmu Komputer;

MOTTO

“Belajar hidup untuk sebuah kehidupan”

*“Walikulli Ummatin Ajalun Fa-Idzaa Jaa-A Ajaluhum Laa Yasta'khiruuna
Saa'atan Walaa Yastaqdimuun”*

~ QS. Al-A'raaf:34 ~

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhamat Abdul Rohim

NIM : 142410101010

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Implementasi Ekstraksi Web (*Web Scraping*) Pada Situs Berita Menggunakan Metode Ekspresi Reguler” adalah benar-benar hasil karya sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya, belum pernah diajukan pada institusi mana pun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa adanya tekanan dan paksaan dari pihak mana saja serta bersedia mendapat sanksi akademik jika di kemudian hari pernyataan ini tidak benar.

Jember, Juli 2018

Yang menyatakan,

Muhamat Abdul Rohim

NIM 142410101010

SKRIPSI

**IMPLEMENTASI EKSTRAKSI WEB (*WEB SCRAPING*) PADA SITUS
BERITA MENGGUNAKAN METODE EKSPRESI REGULER**

Oleh :

Muhamat Abdul Rohim

NIM 142410101010

Pembimbing :

Dosen Pembimbing Utama : Achmad Maududie, ST, M.Sc.

Dosen Pembimbing Pendamping : Windi Eka Yulia Retnani, S.Kom.,M.T

PENGESAHAN PEMBIMBING

Skripsi berjudul “Implementasi Ekstraksi Web (*Web Scraping*) Pada Situs Berita Menggunakan Metode Ekspresi Reguler”, telah diuji dan disahkan pada:

hari, tanggal : Senin, 16 Juli 2018

tempat : Program Studi Sistem Informasi Universitas Jember

Disetujui oleh:

Pembimbing I,

Pembimbing II,

Achmad Maududie, ST, M.Sc
NIP. 197004221995121001

Windi Eka Yulia R, S.Kom., MT
NIP. 198403052010122002

PENGESAHAN

Skripsi berjudul “Implementasi Ekstraksi Web (*Web Scraping*) Pada Situs Berita Menggunakan Metode Ekspresi Reguler”, telah diuji dan disahkan pada:

hari, tanggal : Senin, 16 Juli 2018

tempat : Program Studi Sistem Informasi Universitas Jember

Tim Penguji,

Penguji I,

Penguji II,

Nelly Oktavia A, S.Si., MT.
NIP.198410242009122008

Fajrin Nurman Arifin, S.T., M.Eng.
NIP. 198511282015041002

Mengesahkan

Dekan Fakultas Ilmu Komputer,

Prof. Drs. Slamini, M.Comp.Sc.,Ph.D
NIP. 19670420 1992011001

RINGKASAN

Implementasi Ekstraksi Web (*Web Scraping*) pada Situs Berita Menggunakan Metode Ekspresi Reguler; Muhamat Abdul Rohim, 142410101010; 2018, 222 HALAMAN; Program Studi Sistem Informasi Universitas Jember.

Situs berita, sebagai salah satu penyedia informasi, merupakan web yang sangat sering dikunjungi. Berdasarkan *Alexa Traffic Rank* (ATR), ada tiga situs berita yang termasuk sebagai 10 situs dengan akses tertinggi di Indonesia, yaitu www.detik.com, www.tribunnews.com, dan www.liputan6.com. Sebagai situs berita, ketiga situs web tersebut selalu berusaha menyajikan berita terbaru. Tidak jarang ketiga situs tersebut memuat berita dengan topik yang serupa.

Dari hasil observasi awal yang dilakukan peneliti selama 7 hari berturut-turut terungkap bahwa berita dengan judul yang mengandung kata kunci “mudik” dimuat sebanyak 499 dalam situs www.detik.com, 423 dalam situs www.tribunnews.com dan 174 berita dalam situs www.liputan6.com. Ketika pengguna membutuhkan data dari ketiga situs itu, pengguna harus melakukan proses pengumpulan data secara manual yaitu dengan mengunjungi setiap situs berita menggunakan aplikasi *browser* dan melakukan proses *copy* dan *paste* untuk setiap halaman yang memuat berita yang dimaksud. Proses pengumpulan data seperti ini membutuhkan usaha yang cukup besar dan waktu yang relatif lama, alternatif lain yang bisa dilakukan adalah dengan menggunakan teknik *web scraping*.

Teknik *web scraping* adalah teknik pengambilan data pada file html secara otomatis berdasarkan pola *tag* html yang menyusunnya. Teknik ini dapat dilakukan dengan memanfaatkan pola *regular expression* di dalamnya. Berdasarkan penjelasan sebelumnya, peneliti mencoba membangun sebuah aplikasi yang didasarkan pada teknik *web scraping* menggunakan pola *regular expression* untuk mengekstrak data berita dalam laman situs berita secara otomatis dan disimpan dalam sebuah basis data untuk kebutuhan analisis berita. Pola *Regex* yang dibangun akan didasarkan pada struktur *tag* html yang digunakan pada masing-masing situs, sehingga akan ditemukan perbedaan pola dari masing-masing situs berita yang diinginkan.

Hasil yang ditemukan peneliti setelah melakukan penelitian adalah ditemukannya perbedaan pola penulisan halaman pada masing-masing situs. Perbedaan pola penulisan tersebut mengakibatkan perbedaan pola *regex* yang digunakan untuk melakukan ekstraksi. Perbedaan ini terletak pada jenis *tag* dan atribut serta *value* yang digunakan. Perbedaan pola *regex* yang ditemukan ini telah diimplementasikan pada aplikasi yang dibangun. Proses uji akurasi dilakukan pada tanggal 04 Juli 2018 pukul 08.56 - 08.57 WIB menggunakan tiga metode pengujian, yaitu *precision*, *recall*, dan *f-measure* untuk mengukur tingkat akurasi aplikasi yang dibangun. Hasil uji akurasi terhadap aplikasi menunjukkan bahwa ketiga situs memiliki akurasi sebesar 100%.

PRAKATA

Puji syukur ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Implementasi Ekstraksi Web (*Web Scraping*) Pada Situs Berita Menggunakan Metode Ekspresi Reguler”. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Sistem Informasi Universitas Jember.

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Prof. Drs. Slamin, M.Comp.Sc., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Jember;
2. Achmad Maududie, ST, M.Sc selaku Dosen Pembimbing Utama dan Windi Eka Yulia R, S.Kom., MT selaku Dosen Pembimbing Pendamping yang telah meluangkan waktu, pikiran, dan perhatian dalam penulisan skripsi;
3. Prof. Dr. Saiful Bukhori, S.T., M.Kom. selaku Dosen Pembimbing Akademik (DPA), yang telah mendampingi penulis sebagai mahasiswa.
4. Seluruh Bapak dan Ibu dosen beserta staf karyawan di Program Studi Sistem Informasi Universitas Jember;
5. Ayahanda Muki, Ibunda Rohmani dan nenek saya Alm. Muntini yang selalu mendukung, mendoakan penulis;
6. Sahabat – sahabat Tim ARAY, Angga Dwi Hariadi, Alam Ardianto, dan Yovanka Adam R yang selalu memberikan solusi tidak masuk akal terhadap masalah yang dihadapi penulis.
7. Sahabat-sahabat satu kontrakan saya yang selalu menemani dan memberikan semangat serta bertanya kapan sidang, Muzammil, Moh. Thoif Khairi, dan M. Alviansyah.
8. Fajar Rizky yang telah membantu penulis saat proses penulisan jurnal.
9. Teman-teman seperjuangan SENSATION angkatan 2014.

10. Keluarga besar asisten Laboratorium Rekayasa Perangkat Lunak tahun 2016-2017 dan tahun 2017-2018;
11. Semua mahasiswa Fakultas Ilmu Komputer yang telah menjadi keluarga kecil bagi penulis selama menempuh pendidikan S1;
12. Semua pihak yang tidak dapat disebutkan satu-persatu.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna, oleh sebab itu penulis mengharapkan adanya masukan yang bersifat membangun dari semua pihak. Penulis berharap skripsi ini dapat bermanfaat bagi semua pihak.

Jember, 12 Juli 2018

Penulis

DAFAR ISI

PERSEMBAHAN	ii
MOTTO.....	iii
PERNYATAAN.....	iv
PENGESAHAN PEMBIMBING.....	vi
PENGESAHAN	vii
RINGKASAN	viii
PRAKATA.....	ix
DAFAR ISI.....	xi
DAFTAR GAMBAR	xviii
DAFTAR TABEL.....	xxiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Situs Berita	6
2.2 <i>Document Object Model-Hyperteks Markup Language (DOM-HTML)</i>	7
2.3 <i>Web Scraping</i>	7
2.4 <i>Ekspresi Reguler</i>	9
2.5 <i>Software Development Life Cycle (SDLC) Waterfall</i>	10
BAB 3 METODOLOGI PENELITIAN.....	12
3.1 Jenis Penelitian	12
3.2 Waktu Penelitian.....	12
3.3 Gambaran Aplikasi	12
3.3.1. Scraper.....	13
3.3.2. <i>Monitoring</i>	14
3.4 Tahapan Penelitian.....	15

3.4.1.	Analisis Kebutuhan Pembentukan Pola <i>Regex</i>	16
3.4.2.	Analisa Kebutuhan Aplikasi.....	19
3.4.3.	Desain Aplikasi	19
3.4.4.	Implementasi <i>Web Scraping</i>	20
3.4.5.	Uji Akurasi	21
BAB 4 PENGEMBANGAN APLIKASI.....		23
4.1.	Analisis Struktur Situs Berita	23
4.1.1.	Observasi Penulisan <i>Tautan</i> Berita	23
4.1.2.	Observasi Penulisan Judul Berita.....	26
4.1.3.	Observasi Penulisan Isi Berita	28
4.1.4.	Observasi Penulisan Penulis Berita.....	34
4.1.5.	Observasi Penulisan Tanggal Publikasi Berita.....	36
4.1.6.	Observasi Penulisan <i>Tautan</i> Berita Berupa Video.....	39
4.2.	Pembentukan Pola Ekspresi Reguler.....	42
4.2.1.	Pola <i>Regex Tautan</i> Berita	42
4.2.2.	Pola <i>Regex</i> Judul Berita	43
4.2.3.	Pola <i>Regex</i> Isi Berita	44
4.2.4.	Pola <i>Regex</i> Penulis Berita	48
4.2.5.	Pola <i>Regex</i> Tanggal Publikasi Berita	50
4.2.6.	Pola <i>Regex Tautan</i> Berita Berupa Video	51
4.3.	Analisa Kebutuhan Aplikasi	52
4.3.1.	Kebutuhan Fungsional	52
4.3.2.	Kebutuhan Non-fungsional	53
4.4.	Desain Aplikasi.....	54
4.4.1.	<i>Bussiness Process</i>	54
4.4.2.	Usecase Diagram.....	57
4.4.3.	Usecase Skenario	60
4.4.4.	Sequence Diagram	68
4.4.5.	Activity Diagram.....	75
4.4.6.	Entity Relationship Diagram (ERD)	85
4.5.	Implementasi <i>Web Scraping</i>	85
4.5.1.	<i>Scraper</i>	85

4.5.2. <i>Monitoring</i>	89
4.6. Uji Akurasi	93
BAB 5 HASIL DAN PEMBAHASAN.....	95
5.1. Hasil analisis struktur halaman.....	95
5.1.1. Pola penulisan tautan berita	95
5.1.2. Pola Penulisan Judul Berita.....	96
5.1.3. Pola penulisan isi berita	96
5.1.4. Pola Penulisan Penulis/Sumber Berita	98
5.1.5. Pola penulisan tanggal publikasi berita.....	98
5.1.6. Pola Penulisan Tautan Video	99
5.2. Hasil Pembentukan Pola <i>Regex</i>	99
5.2.1. Pola <i>Regex</i> yang Ditemukan	100
5.2.2. Perbedaan Pola <i>Regex</i>	107
5.3. Hasil implementasi aplikasi.....	109
5.3.1. <i>Scraper</i>	109
5.3.2. <i>Monitoring</i>	110
5.4. Hasil Uji Akurasi	122
5.4.1. Detik.....	122
5.4.2. Liputan6	124
5.4.3. Tribunnews.....	125
BAB 6 PENUTUP	126
6.1. Kesimpulan.....	126
6.2. Saran.....	128
DAFTAR PUSTAKA	129
LAMPIRAN.....	131
Lampiran A. Skenario	131
A.1. Melihat Beranda	131
A.2. Mendaftar sebagai Anggota.....	131
A.3. Melihat Statistik Ekstraksi	136
A.4. Mencari Detail Statistik.....	137
A.5. Melihat Daftar Situs Berita	138
A.6. Masuk Aplikasi	138

A.7.	Melihat Daftar Berita	139
A.8.	Mencari Daftar Berita	140
A.9.	Memfilter Daftar Berita.....	141
A.10.	Melihat Detail Berita.....	142
A.11.	Melihat Profil	142
A.12.	Mengubah Profil	143
A.13.	Keluar Aplikasi.....	144
A.14.	Melihat Daftar Pengajuan Anggota.....	144
A.15.	Mencari Daftar Pengajuan.....	145
A.16.	Melihat Detai Pengajuan	145
A.17.	Menyetujui Pengajuan Anggota	146
A.18.	Menolak Pengajuan Anggota	146
A.19.	Melihat Daftar Anggota.....	147
A.20.	Memfilter Daftar Anggota.....	148
A.21.	Melihat Detail Anggota	148
A.22.	Menonaktifkan Keanggotaan	149
A.23.	Mengaktifkan Keanggotaan	149
A.24.	Menghapus Keanggotaan	149
A.25.	Melihat Daftar Jenis Berita	150
A.26.	Menambahkan Jenis Berita Baru	150
A.27.	Mengubah Jenis Berita.....	151
A.28.	Menghapus Jenis Berita	152
A.29.	Melihat Daftar Pola Ekspresi Regular.....	152
A.30.	Menambahkan Pola Eskpresi Regular	153
A.31.	Mengubah Pola Ekspresi Regular	153
A.32.	Menghapus Pola Ekspresi Reular.....	154
A.33.	Melihat Daftar Pola Khusus	155
A.34.	Menambahkan Pola Khusus.....	155
A.35.	Mengubah Pola Khusus	156
A.36.	Menghapus Pola Khusus	157
A.37.	Melihat Daftar Filter HTML	157
A.38.	Menambahkan Filter HTML Baru	157

A.39.	Mengubah Filter HTML.....	158
A.40.	Menghapus Filter HTML.....	159
A.41.	Mencari judul berita.....	159
Lampiran B.	Sequence Diagram.....	160
B.1.	Melihat Beranda.....	160
B.2.	Mendaftar sebagai Anggota.....	160
B.3.	Melihat Statistik Ekstraksi.....	161
B.4.	Mencari Detail Statistik.....	161
B.5.	Melihat Daftar Situs Berita.....	161
B.6.	Masuk Aplikasi.....	162
B.7.	Melihat Daftar Berita.....	162
B.8.	Mencari Daftar Berita.....	163
B.9.	Memfilter Daftar Berita.....	163
B.10.	Melihat Detail Berita.....	164
B.11.	Melihat Profil.....	164
B.12.	Mengubah Profil.....	165
B.13.	Keluar Aplikasi.....	166
B.14.	Melihat Daftar Pengajuan Anggota.....	166
B.15.	Mencari Daftar Pengajuan.....	166
B.16.	Melihat Detail Pengajuan.....	167
B.17.	Menyetujui Pengajuan Anggota.....	167
B.18.	Menolak Pengajuan Anggota.....	167
B.19.	Melihat Daftar Anggota.....	168
B.20.	Memfilter Daftar Anggota.....	168
B.21.	Melihat Detail Anggota.....	169
B.22.	Menonaktifkan Keanggotaan.....	169
B.23.	Mengaktifkan Keanggotaan.....	170
B.24.	Menghapus Keanggotaan.....	170
B.25.	Melihat Daftar Jenis Berita.....	171
B.26.	Menambahkan Jenis Berita Baru.....	171
B.27.	Mengubah Jenis Berita.....	172
B.28.	Menghapus Jenis Berita.....	172

B.29.	Melihat Daftar Pola Ekspresi Regular.....	173
B.30.	Menambahkan Pola Eskpresi Regular	173
B.31.	Mengubah Pola Ekspresi Regular	174
B.32.	Menghapus Pola Ekspresi Regular.....	174
B.33.	Melihat Daftar Pola Khusus	175
B.34.	Menambahkan Daftar Pola Khusus.....	175
B.35.	Mengubah Pola Khusus	176
B.36.	Menghapus Pola Khusus	176
B.37.	Melihat Daftar Filter HTML	177
B.38.	Manambahkan Filter HTML Baru	177
B.39.	Mengubah Filter HTML.....	178
B.40.	Menghapus Filter HTML	178
B.41.	Mencari Judul Berita	179
Lampiran C.	Activity Diagram.....	179
C.1.	Melihat Beranda	179
C.2.	Mendaftar sebagai Anggota.....	180
C.3.	Melihat Statistik Ekstraksi	181
C.4.	Mencari Detai Statistik.....	181
C.5.	Melihat Daftar Situs Berita	182
C.6.	Masuk Aplikasi	182
C.7.	Melihat Daftar Berita	183
C.8.	Mencari Daftar Berita	183
C.9.	Memfilter Daftar Berita.....	184
C.10.	Melihat Detail Berita.....	184
C.11.	Melihat Profil	184
C.12.	Mengubah Profil	185
C.13.	Keluar Aplikasi.....	185
C.14.	Melihat Daftar Pengajuan Anggota.....	186
C.15.	Mencari Daftar Pengajuan.....	186
C.16.	Melihat Detai Pengajuan	186
C.17.	Menyetujui Pengajuan Anggota	187
C.18.	Menolak Pengajuan Anggota	187

C.19.	Melihat Daftar Anggota.....	187
C.20.	Memfilter Daftar Anggota.....	188
C.21.	Melihat Detail Anggota.....	188
C.22.	Menonaktifkan Keanggotaan.....	188
C.23.	Mengaktifkan Keanggotaan.....	189
C.24.	Menghapus Keanggotaan.....	189
C.25.	Melihat Daftar Jenis Berita.....	190
C.26.	Menambahkan Jenis Berita Baru.....	190
C.27.	Mengubah Jenis Berita.....	190
C.28.	Menghapus Jenis Berita.....	191
C.29.	Melihat Daftar Pola Ekspresi Regular.....	191
C.30.	Menambahkan Pola Ekspresi Regular.....	191
C.31.	Mengubah Pola Ekspresi Regular.....	192
C.32.	Menghapus Pola Ekspresi Regular.....	192
C.33.	Melihat Daftar Pola Khusus.....	193
C.34.	Menambahkan Daftar Pola Khusus.....	193
C.35.	Mengubah Pola Khusus.....	193
C.36.	Menghapus Pola Khusus.....	194
C.37.	Melihat Daftar Filter HTML.....	194
C.38.	Menambahkan Filter HTML Baru.....	194
C.39.	Mengubah Filter HTML.....	195
C.40.	Menghapus Filter HTML.....	195
C.41.	Mencari Judul Berita.....	195

DAFTAR GAMBAR

Gambar 2. 1 Langkah-langkah Web Scraping	9
Gambar 2. 2 Model Waterfall (Youssef Bassil 2012).....	11
Gambar 3. 1 Alur Kerja Aplikasi.....	14
Gambar 3. 2 Proses bisnis sub aplikasi monitoring	15
Gambar 3. 3 Tahapan Penelitian	16
Gambar 3. 4 Struktur halaman indeks berita Detik.com.....	17
Gambar 3. 5 Struktur halaman berita Detik.com	17
Gambar 3. 6 Struktur Penulisan Tautan Berita Detik.com.....	19
Gambar 4. 1 Struktur halaman indeks www.detik.com	23
Gambar 4. 2 Struktur halaman indeks www.liputan6.com	24
Gambar 4. 3 Struktur halaman indeks www.tribunnews.com.....	25
Gambar 4. 4 Struktur halaman berita video	26
Gambar 4. 5 Struktur html halaman berita narasi/tekstual.....	26
Gambar 4. 6 Struktur halaman berita liputan6.com	27
Gambar 4. 7 Struktur html halaman berita tribunnews.com	28
Gambar 4. 8 Struktur html halaman berita video detik.com.....	28
Gambar 4. 9 Struktur html halaman berita narasi/tekstual.....	29
Gambar 4. 10 Struktur halaman berita situs liputan6.com.....	31
Gambar 4. 11 Struktur halaman berita www.tribunnews.com	32
Gambar 4. 12 Struktur html halaman berita video detik.com.....	34
Gambar 4. 13 Struktur html halaman berita narasi/tekstual detik.com.....	35
Gambar 4. 14 Struktur html halaman berita liputan6.com.....	35
Gambar 4. 15 Struktur Penulisan Nama Penulis pada www.tribunnews.com	36
Gambar 4. 16 Struktur html halaman berita video detik.com	37
Gambar 4. 17 struktur html halaman berita narasi/tekstual detik.com	37
Gambar 4. 18 Struktur html halaman berita liputan6.com.....	38

Gambar 4. 19 Struktur Penulisan Tanggal Publikasi www.tribunnews.com	39
Gambar 4. 20 Struktur penulisan tautan berita video detik.com.....	40
Gambar 4. 21 Struktur Penulisan Tautan Video Berita Video www.liputan6.com	40
Gambar 4. 22 Struktur Penulisan Tautan Video Berita Narasi www.liputan6.com	41
Gambar 4. 23 Struktur Penulisan Tautan Video pada www.tribunnews.com	41
Gambar 4. 24 <i>Bussiness Process</i>	55
Gambar 4. 25 BPMN pengelolaan anggota.....	56
Gambar 4. 26 BPMN pengelolaan pola regex	56
Gambar 4. 27 Usecase Diagram.....	57
Gambar 4. 28 Sequence Diagram melakukan ekstraksi Situs Berita	69
Gambar 4. 29 Activity Melakukan Ekstraksi Situs Berita	76
Gambar 4. 30 Class Diagram Scrapper	84
Gambar 4. 31 Class Diagram Site Monitoring.....	84
Gambar 4. 32 <i>Entiy Relationship Diagram</i>	85
Gambar 4. 33 Class ConnectioClass	86
Gambar 4. 34 Class ScrapperMaster	86
Gambar 4. 35 Class DetikScrapper	87
Gambar 4. 36 Class Liputan6Scrapper	87
Gambar 4. 37 Class TribunScrapper	88
Gambar 4. 38 Class Program	88
Gambar 4. 39 Class connectionClass	89
Gambar 4. 40 Class MainGuest dan relasinya	90
Gambar 4. 41 Class MainMember dan relasinya	91
Gambar 4. 42 Class MainAdmin dan relasinya	91
Gambar 4. 43 Class SiteController.....	92
Gambar 4. 44 Class NewsController.....	92
Gambar 5. 1 Tampilan Scrapper.....	110
Gambar 5. 2 Tampilan beranda	111
Gambar 5. 3 Tampilan pendaftaran anggota	111

Gambar 5. 4 Tampilan statistik ekstraksi	112
Gambar 5. 5 Tampilan daftar situs berita	112
Gambar 5. 6 Tampilan daftar situs berita	113
Gambar 5. 7 Tampilan daftar berita	113
Gambar 5. 8 Tampilan detail berita	114
Gambar 5. 9 Tampilan Profil	114
Gambar 5. 10 Tampilan mengubah profil	115
Gambar 5. 11 Tampilan filter berita	115
Gambar 5. 12 Tampilan daftar pengajuan keanggotaan	116
Gambar 5. 13 Tampilan detail pengajuan	116
Gambar 5. 14 Tampilan daftar anggota	117
Gambar 5. 15 Tampilan detail anggota	117
Gambar 5. 16 Tampilan daftar jenis berita	118
Gambar 5. 17 Tampilan daftar pola <i>regex</i>	118
Gambar 5. 18 Tampilan daftar Pola Khusus	119
Gambar 5. 19 Tampilan daftar filter html	119
Gambar B. 1 Sequence diagram melihat beranda	160
Gambar B. 2 Sequence diagram mendaftar sebagai anggota	160
Gambar B. 3 Sequence diagram melihat statistik ekstraksi	161
Gambar B. 4 Sequence diagram detail statistik	161
Gambar B. 5 Sequence diagram melihat daftar situs berita	161
Gambar B. 6 Sequence diagram masuk aplikasi	162
Gambar B. 7 Sequence diagram melihat daftar berita	162
Gambar B. 8 Sequence diagram mencari daftar berita	163
Gambar B. 9 Sequence diagram memfilter daftar berita	163
Gambar B. 10 Sequence diagram melihat detail berita	164
Gambar B. 11 Sequence diagram melihat profil	164
Gambar B. 12 Sequence diagram mengubah foto profil (admin)	165
Gambar B. 13 Sequence diagram mengubah foto profil (anggota)	165
Gambar B. 14 Sequence diagram keluar aplikasi	166

Gambar B. 15 Sequence diagram melihat daftar anggota.....	166
Gambar B. 16 Sequence diagram mencari daftar pengajuan	166
Gambar B. 17 Sequence diagram melihat detail pengajuan	167
Gambar B. 18 Sequence diagram menyetujui pengajuan anggota.....	167
Gambar B. 19 Sequence diagram menolak pengajuan anggota.....	167
Gambar B. 20 Sequence diagram melihat daftar anggota.....	168
Gambar B. 21 Sequence diagram memfilter daftar anggota	168
Gambar B. 22 Sequence diagram melihat detail anggota	169
Gambar B. 23 Sequence diagram menonaktifkan keanggotaan.....	169
Gambar B. 24 Sequence diagram mengaktifkan keanggotaan.....	170
Gambar B. 25 Sequence diagram menghapus keanggotaan	170
Gambar B. 26 Sequence diagram melihat jenis berita	171
Gambar B. 27 Sequence diagram menambahkan jenis berita.....	171
Gambar B. 28 Sequence diagram mengubah jenis berita	172
Gambar B. 29 Sequence diagram menghapus jenis berita.....	172
Gambar B. 30 Sequence diagram melihat daftar pola regex.....	173
Gambar B. 31 Sequence diagram menambahkan pola regex.....	173
Gambar B. 32 Sequence diagram mengubah pola regex	174
Gambar B. 33 Sequence diagram menghapus pola regex.....	174
Gambar B. 34 Sequence diagram melihat daftar pola khusus	175
Gambar B. 35 Sequence diagram menambahkan pola khusus	175
Gambar B. 36 Sequence diagram mengubah pola khusus	176
Gambar B. 37 Sequence diagram menghapus pola khusus.....	176
Gambar B. 38 Sequence diagram melihat filter html.....	177
Gambar B. 39 Sequence diagram menambahkan filter html	177
Gambar B. 40 Sequence diagram mengubah filter html	178
Gambar B. 41 Sequence diagram menghapus filter html	178
Gambar B. 42 Sequence diagram mencari judul berita.....	179
Gambar C. 1 Activity Diagram Melihat Beranda.....	179
Gambar C. 2 Activity Diagram Mendaftar Anggota	180

Gambar C. 3 Activity Diagram Melihat Statistik.....	181
Gambar C. 4 Activity Diagram Mencari Detail Statistik	181
Gambar C. 5 Activity Diagram Melihat Daftar Situs Berita.....	182
Gambar C. 6 Activity Diagram Masuk Aplikasi	182
Gambar C. 7 Activity Diagram Melihat Daftar Berita.....	183
Gambar C. 8 Activity Diagram Mencari Daftar Berita	183
Gambar C. 9 Activity Diagram Memfilter Daftar Berita	184
Gambar C. 10 Activity Diagram Melihat Detail Berita	184
Gambar C. 11 Activity Diagram Melihat Profil.....	184
Gambar C. 12 Activity Diagram Mengubah Profil (admin)	185
Gambar C. 13 Activity Diagram Mengubah Profil (anggota).....	185
Gambar C. 14 Activity Diagram Keluar Aplikasi	185
Gambar C. 15 Activity Diagram Melihat Daftar Pengajuan Anggota	186
Gambar C. 16 Activity Diagram Mencari Daftar Pengajuan	186
Gambar C. 17 Activity Diagram Melihat Detail Pengajuan	186
Gambar C. 18 Activity Diagram Menyetujui Pengajuan	187
Gambar C. 19 Activity Diagram Menolak Pengajuan Anggota.....	187
Gambar C. 20 Activity Diagram Melihat Daftar Anggota	187
Gambar C. 21 Activity Diagram Memfilter Daftar Anggota	188
Gambar C. 22 Activity Diagram Melihat Detail Anggota.....	188
Gambar C. 23 Activity Diagram Menonaktifkan Keanggotaan.....	188
Gambar C. 24 Activity Diagram Mengaktifkan Keanggotaan.....	189
Gambar C. 25 Activity Diagram Menghapus Keanggotaan.....	189
Gambar C. 26 Activity Diagram Melihat Daftar Jenis Berita.....	190
Gambar C. 27 Activity Diagram Menambah Jenis Berita.....	190
Gambar C. 28 Activity Diagram Mengubah Jenis Berita	190
Gambar C. 29 Activity Diagram Menghapus Jenis Berita	191
Gambar C. 30 Activity Diagram Melihat Pola Regex.....	191
Gambar C. 31 Activity Diagram Menambah Pola Regex	191
Gambar C. 32 Activity Diagram Mengubah Pola Regex	192
Gambar C. 33 Activity Diagram Menghapus Pola Regex	192

Gambar C. 34 Activity Diagram Melihat Daftar Pola Khusus.....	193
Gambar C. 35 Activity Diagram Menambah Pola Khusus	193
Gambar C. 36 Activity Diagram Mengubah Pola Khusus	193
Gambar C. 37 Activity Diagram Menghapus Pola Khusus.....	194
Gambar C. 38 Activity Diagram Melihat Daftar Filter HTML.....	194
Gambar C. 39 Activity Diagram Menambah Filter HTML	194
Gambar C. 40 Activity Diagram Mengubah Filter HTML	195
Gambar C. 41 Menghapus Filter HTML Menghapus Filter HTML	195
Gambar C. 42 Activity Diagram Mencari Judul Berita.....	195

DAFTAR TABEL

Tabel 2. 1 Pola Ekspresi Reguler	9
Tabel 4. 1 Deskripsi Aktor	58
Tabel 4. 2 Deskripsi Usecase	58
Tabel 4. 3 Usecase skenario melakukan ekstraksi situs berita	60
Tabel 5. 1 Hasil analisis struktur penulisan tautan berita.....	95
Tabel 5. 2 Hasil analisis struktur penulisan judul berita	96
Tabel 5. 3 Hasil analisis struktu isi berita	97
Tabel 5. 4 Pola penulisan sisipan non konten	97
Tabel 5. 5 Hasil analisis struktur penulisan penulis berita	98
Tabel 5. 6 Hasil analisis struktur penulisan tanggal berita.....	98
Tabel 5. 7 Hasil analisis struktur penulisan tautan video berita.....	99
Tabel 5. 8 Pola regex tautan berita	100
Tabel 5. 9 Pola regex judul berita.....	101
Tabel 5. 10 Pola regex situs Detik (Narasi).....	102
Tabel 5. 11 Pola regex situs Detik (Video).....	102
Tabel 5. 12 Pola regex situs Liputan6	103
Tabel 5. 13 Pola regex situs Tribunnews.....	104
Tabel 5. 14 Filter html standar	105
Tabel 5. 15 Filter khusus masing-masing situs	105
Tabel 5. 16 Pola regex penulis berita	105
Tabel 5. 17 Pola regex tanggal publikasi berita	106
Tabel 5. 18 Pola regex tautan video berita	107
Tabel 5. 19 Perbedaan antar pola masing-masing situs berita	107
Tabel 5. 20 Hasil pengujian black box pada sub aplikasi monitoring.....	120
Tabel 5. 21 Data uji akurasi Detik.....	123
Tabel 5. 22 Data uji akurasi Liputan6	124

Tabel 5. 23 Data uji akurasi Liputan6	125
Tabel 6. 1 Pola regex yang ditemukan	126
Tabel A. 1 Skenario melihat beranda	131
Tabel A. 2 Skenario mendaftar sebagai anggota	131
Tabel A. 3 Skenario melihat detail statistik.....	136
Tabel A. 4 Skenario mencari detail statistik.....	137
Tabel A. 5 Skenario melihat daftar situs berita	138
Tabel A. 6 Skenario masuk aplikasi	138
Tabel A. 7 Skenario melihat daftar berita.....	139
Tabel A. 8 Skenario mencari daftar berita.....	140
Tabel A. 9 Skenario memfilter daftar berita.....	141
Tabel A. 10 Skenario melihat detail berita	142
Tabel A. 11 Skenario melihat profil	142
Tabel A. 12 Skenario mengubah profil anggota.....	143
Tabel A. 13 Skenario mengubah profil admin	143
Tabel A. 14 Skenario keluar aplikasi.....	144
Tabel A. 15 Skenario melihat daftar pengajuan anggota.....	144
Tabel A. 16 Skenario mencari daftar pengajuan	145
Tabel A. 17 Skenario melihat detail pengajuan.....	145
Tabel A. 18 Skenario menyetujui keanggotaan.....	146
Tabel A. 19 Skenario menolak pengajuan anggota	146
Tabel A. 20 Skenario melihat daftar anggota	147
Tabel A. 21 Skenario memfilter daftar anggota	148
Tabel A. 22 Skenario melihat detail anggota	148
Tabel A. 23 Skenario menonaktifkan keanggotaan.....	149
Tabel A. 24 Skenario mengaktifkan keanggotaan.....	149
Tabel A. 25 Skenario menghapus keanggotaan.....	149
Tabel A. 26 Skenario melihat daftar jenis berita	150

Tabel A. 27 Skenario menambah jenis berita baru.....	150
Tabel A. 28 Skenario mengubah jenis berita.....	151
Tabel A. 29 Skenario menghapus jenis berita	152
Tabel A. 30 Skenario melihat daftar pola regex	152
Tabel A. 31 Skenario menambah pola regex.....	153
Tabel A. 32 Skenario mengubah pola regex.....	153
Tabel A. 33 Skenario menghapus pola regex	154
Tabel A. 34 Skenario melihat daftar Pola Khusus.....	155
Tabel A. 35 Skenario menambahkan pola khusus.....	155
Tabel A. 36 Skenario mengubah pola khusus	156
Tabel A. 37 Skenario menghapus pola khusus.....	157
Tabel A. 38 Skenario melihat daftar filter HTML.....	157
Tabel A. 39 Skenario menambahkan filter html baru	157
Tabel A. 40 Skenario mengubah filter html	158
Tabel A. 41 Skenario menghapus filter html.....	159
Tabel A. 42 Skenario mencari judul berita.....	159

BAB 1 PENDAHULUAN

Bab ini merupakan langkah awal dari penulisan skripsi ini. Bab ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan Sistematika penulisan.

1.1 Latar Belakang

Indonesia merupakan salah satu negara dengan penetrasi pengguna Internet yang tinggi. Berdasarkan hasil survei Asosiasi Pengusaha Jasa Internet Indonesia (APJII) tahun 2017, penetrasi pengguna Internet di Indonesia telah mencapai 143,26 juta jiwa (54,68% dari jumlah penduduk) di tahun 2017. Dari pengguna tersebut, 58,01% di antaranya memanfaatkan Internet untuk membaca berita *entertainment*, 50,48% membaca berita olah raga, 50,26% membaca berita sosial/lingkungan, 41,55% membaca informasi agama, 36,94% membaca berita politik, dan 51,06% mencari informasi tentang kesehatan (APJII 2017).

Situs berita, sebagai salah satu penyedia informasi, merupakan web yang sangat sering dikunjungi. Berdasarkan *Alexa Traffic Rank* (ATR), ada tiga situs berita yang termasuk sebagai 10 situs dengan akses tertinggi di Indonesia, yaitu www.detik.com, www.tribunnews.com, dan www.liputan6.com (ATR 2017). Tingginya tingkat akses ini menunjukkan bahwa ketiga situs web tersebut sering dijadikan rujukan untuk mendapatkan informasi.

Sebagai situs berita, ketiga situs web tersebut (www.detik.com, www.tribunnews.com, dan www.liputan6.com) selalu berusaha menyajikan berita terbaru. Tidak jarang ketiga situs tersebut memuat berita dengan topik yang serupa. Dari hasil observasi awal yang dilakukan peneliti selama 7 hari berturut-turut terungkap bahwa berita dengan judul yang mengandung kata kunci “mudik” dimuat sebanyak 499 dalam situs www.detik.com, 423 dalam situs www.tribunnews.com dan 174 berita dalam situs www.liputan6.com. Pengguna harus mengumpulkan semua berita-berita tersebut dari masing-masing situs untuk dapat digunakan sebagai data analisa kondisi yang menggambarkan situasi mudik.

Proses pengumpulan data berupa berita dari media berbasis daring (*online*)

sering kali menjadi pekerjaan yang sulit untuk keperluan penyediaan data untuk kebutuhan analisa. Permasalahan yang muncul pada pengumpulan berita adalah besarnya jumlah berita yang menjadi kendala, dan tersebarnya berita di berbagai situs juga merupakan kendala tersendiri. Kendala besar dalam proses pengumpulan data dari situs berita adalah tidak disediakannya layanan berbagi data (misal dalam bentuk *web service* ataupun *web API*) oleh situs berita yang bersangkutan. Pengguna, dalam hal ini, harus melakukan proses pengumpulan data secara manual yaitu dengan mengunjungi setiap situs berita menggunakan aplikasi *browser* dan melakukan proses *copy* dan *paste* untuk setiap laman yang memuat berita yang dimaksud. Proses pengumpulan data seperti ini membutuhkan usaha yang cukup besar dan waktu yang relatif lama. Pengumpulan data dari berbagai sumber seperti ini lebih dikenal dengan istilah *data scraping*.

Data scraping adalah proses ekstraksi data dari *file* elektronik menggunakan program komputer. Apabila *file* elektronik tersebut berupa *file HTML*, yang ada di Internet, maka proses ekstraksi data berdasarkan pola-pola *file html* dan proses ini dikenal dengan *Web Scraping* (Haddaway 2015). Teknik *Web Scraping* ini telah dikembangkan oleh sejumlah penelitian, sebagai contoh Muliantara (2009) dan Utomo (2013) membentuk pola berdasarkan ekspresi reguler (*Regex*) untuk mendapatkan informasi dari laman web. Hasil kedua penelitian tersebut menunjukkan bahwa *Regex* dapat menangani penulisan alamat *email* pada konten *wordpress* untuk dijadikan gambar (Muliantara 2009) dan teknik *web scraping* menggunakan *Regex* juga dapat menangani pencarian dokumen berdasarkan tren pencarian di *google* (Utomo 2013).

Berdasarkan penjelasan sebelumnya, peneliti mencoba membangun sebuah aplikasi yang mampu membantu mengumpulkan data dari laman web yang tidak menyediakan layanan berbagi data. Aplikasi ini didasarkan pada teknik *web scraping* untuk mengekstrak data berita dalam laman situs berita secara otomatis dan disimpan dalam sebuah basis data untuk kebutuhan analisis berita. Dalam penelitian ini, pola yang dibangun adalah pola *Regex*.

Pola *Regex* yang dibangun akan didasarkan pada struktur *tag html* yang digunakan pada masing-masing situs, sehingga akan ditemukan perbedaan pola dari

masing-masing situs berita yang diinginkan. Sebagai langkah awal, tiga situs berita yang akan dijadikan objek *scraping* adalah www.detik.com, www.tribunnews.com dan www.liputan6.com, sehingga akan ditemukan bagaimana pola *Regex* dari tiga situs berita tersebut, dan apa perbedaan pola dari masing-masing situs berita.

1.2 Perumusan Masalah

Berdasarkan uraian dalam latar belakang maka rumusan masalah dalam penelitian ini dapat diuraikan sebagai berikut.

1. Bagaimana pola *Regex* untuk melakukan ekstraksi data berita dari masing-masing situs berita yang tidak menyediakan layanan berbagi data?
2. Apa perbedaan pola *Regex* yang ditemukan pada masing-masing situs berita?

1.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini sebagai berikut.

1. Menemukan bagaimana pola *Regex* dari masing-masing situs berita.
2. Menemukan perbedaan pola *Regex* yang ditemukan pada masing-masing situs berita berdasarkan pola penulisan yang digunakan masing-masing situs.

1.4 Batasan Masalah

Beberapa hal yang menjadi batasan masalah pada penelitian ini adalah sebagai berikut.

1. Situs berita yang menjadi objek *scraping* ada tiga, yaitu www.detik.com, www.tribunnews.com dan www.liputan6.com.
2. Teknik *web scraping* dilakukan dengan menggunakan *user-agent mobile*.
3. Aplikasi yang dibangun didesain agar mampu untuk mengekstrak bagian laman web yang dianggap penting saja, yaitu konten berita yang berupa teks dan video.
4. Konten berita yang di ekstrak adalah judul berita, penulis/sumber berita, tanggal publikasi berita, alamat berita, isi berita dan tautan dari berita berupa video.
5. Tahap *maintenance* pada pengembangan aplikasi tidak dilakukan.

6. Akurasi aplikasi yang dikembangkan minimal 90%.
7. Aplikasi dirancang dengan fungsionalitas sebagai berikut:
 - a. Aplikasi mampu melakukan ekstraksi konten berita.
 - b. Aplikasi dapat membedakan akses dari 3 aktor (tamu, anggota, dan admin).
 - c. Aplikasi dapat mengelola pendaftaran anggota.
 - d. Aplikasi dapat menampilkan statistik ekstraksi.
 - e. Aplikasi dapat menampilkan informasi *traffic* masing-masing situs berdasarkan data dari www.alexacom.com.
 - f. Aplikasi dapat menampilkan berita hasil *scraping*.
 - g. Aplikasi dapat mengelola profil pengguna.
 - h. Aplikasi memiliki layanan *API*.
 - i. Aplikasi dapat mengelola jenis berita
 - j. Aplikasi dapat mengelola pola *regex* yang digunakan *scraper*.
 - k. Aplikasi dapat mengelola pola khusus yang dibutuhkan *scraper*.
 - l. Aplikasi dapat mengelola filter html yang digunakan *scraper*.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun sebagai berikut:

1. Pendahuluan
Bab ini menjelaskan tentang latar belakang, perumusan masalah, tujuan, ruang lingkup studi atau batasan masalah dan sistematika penulisan.
2. Tinjauan pustaka
Bab ini menjelaskan tentang materi, informasi, tinjauan pustaka, dan studi terdahulu yang menjadi kerangka pemikiran dalam penelitian.
3. Metodologi penelitian
Bab ini menjelaskan tentang jenis penelitian, tempat dan waktu penelitian, tahapan penelitian, dan gambaran aplikasi yang akan dibangun.
4. Pengembangan aplikasi
Bab ini akan membahas tentang pengembangan aplikasi ekstraksi web (*web scraping*) dengan mengimplementasikan metode ekspresi reguler (*regular*

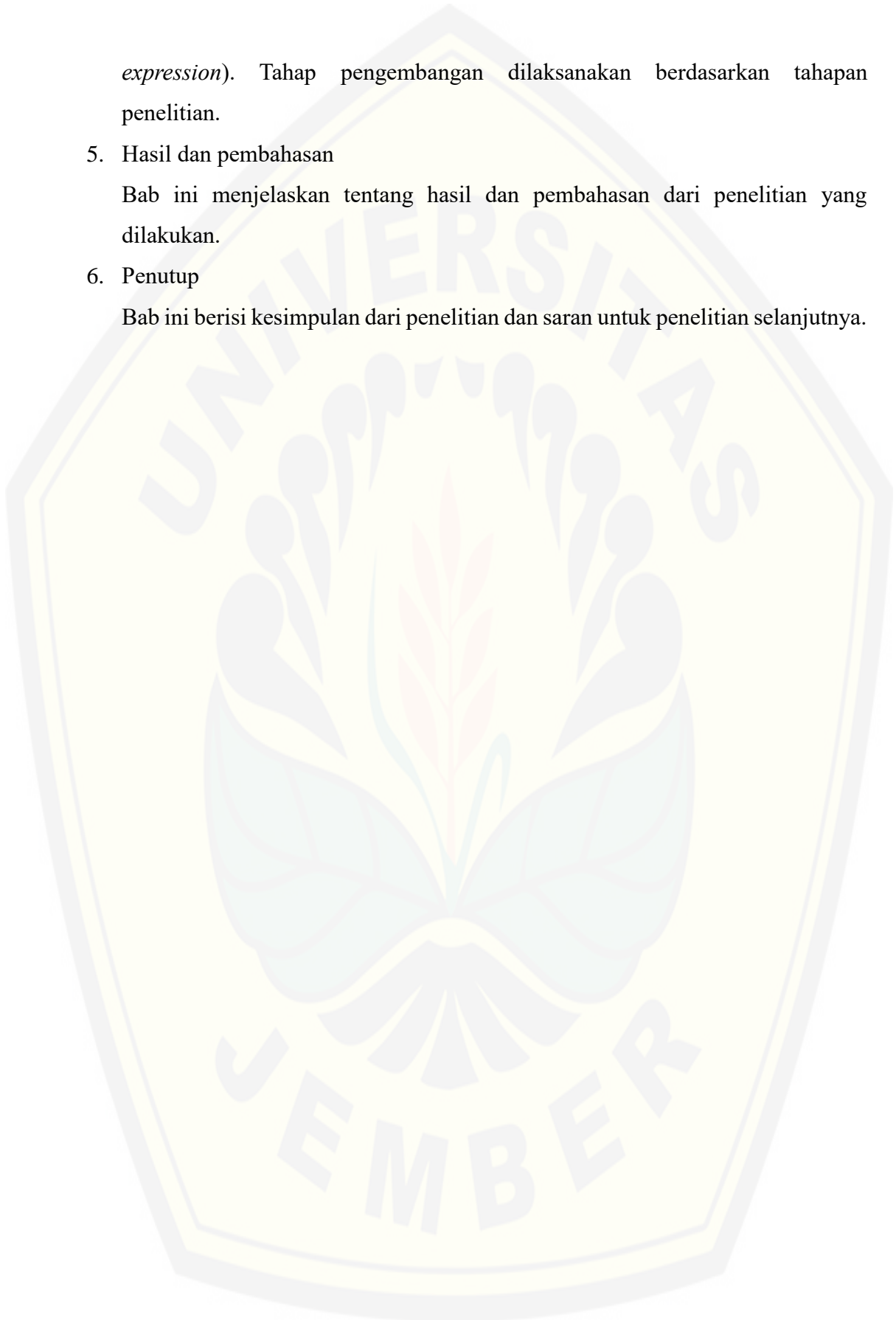
expression). Tahap pengembangan dilaksanakan berdasarkan tahapan penelitian.

5. Hasil dan pembahasan

Bab ini menjelaskan tentang hasil dan pembahasan dari penelitian yang dilakukan.

6. Penutup

Bab ini berisi kesimpulan dari penelitian dan saran untuk penelitian selanjutnya.



BAB 2 TINJAUAN PUSTAKA

Bagian ini memaparkan tinjauan yang berkaitan dengan masalah yang dibahas, kajian teori yang berkaitan dengan masalah, kerangka pemikiran yang merupakan sintesis dari kajian teori yang dikaitkan dengan permasalahan yang dihadapi. Teori-teori ini diambil dari buku, *literature*, jurnal, dan internet.

2.1 Situs Berita

Menurut Yuhefizar *website* adalah keseluruhan halaman-halaman web yang terdapat dari sebuah domain yang mengandung informasi (Prayitno 2015). Sebuah *website* merupakan bagian dari *World Wide Web* (WWW) yang dibuat pada tahun 1989 oleh CERN *Physicist Tim Berners-Lee*. *World Wide Web* mulai dibuka untuk digunakan dengan bebas oleh siapa saja pada 30 April 1993 (Cailliau 2018). Melalui *website*, siapa saja di dunia dapat mengetahui informasi yang ingin disampaikan kapan saja.

Informasi yang disediakan oleh sebuah *website* sangat beragam, berita adalah salah satu informasi yang banyak di muat di berbagai situs, situs yang khusus memuat berita dikenal sebagai situs berita. Berdasarkan *Alexa Traffic Rank* (ATR), terdapat 10 situs dengan akses tertinggi di Indonesia, tiga situs *search engine*, dua situs sosial media, dua situs *online shop*, dan tiga situs berita. Situs berita menjadi situs yang memiliki *traffic* tertinggi setelah *search engine*, hal ini menunjukkan situs berita adalah situs yang diminati untuk di akses oleh masyarakat Indonesia. Tiga situs berita yang menjadi tiga urutan pertama kategori situs berita di www.alexacom untuk negara Indonesia adalah www.detik.com, www.tribunnews.com, dan www.liputan6.com. Berdasarkan ATR dari masing-masing situs berita tersebut, dapat disimpulkan bahwa situs-situs tersebut memiliki *traffic* akses yang tinggi dibandingkan dengan situs berita lain yang ada.

Tingginya tingkat akses ini menunjukkan bahwa ketiga situs tersebut sering dijadikan rujukan untuk mendapatkan informasi. Informasi dalam situs berita ditampilkan dalam halaman-halaman yang berbeda-beda untuk setiap beritanya. Halaman-halaman tersebut dibangun menggunakan *Document Object Model*-

Hyperteks Markup Language (DOM-HTML) sebagai dasar penyusunannya.

2.2 *Document Object Model-Hyperteks Markup Language (DOM-HTML)*

DOM atau *Document Object Model* adalah model standar penulisan sebuah dokumen XML atau HTML. DOM yang digunakan sebagai standar penulisan dokumen HTML dikenal sebagai DOM-HTML. DOM-HTML adalah standar bagaimana untuk mendapatkan, mengubah, menambah, atau menghapus elemen html (Vivensius Mitra and Dr Herry Sujaini 2017). DOM-HTML digunakan untuk menjaga kompatibilitas struktur dokumen html terhadap berbagai aplikasi peramban yang ada. DOM-HTML juga mengatur bagaimana elemen-elemen html disusun agar memenuhi standar yang ada saat ini.

Elemen html adalah kode html yang dimulai dari sebuah *tag* pembuka dan diakhiri dengan *tag* penutup, *tag* itu sendiri adalah kode-kode tertentu yang menjadi pengenalan bahwa kode tersebut adalah kode html dan bisa diterjemahkan oleh browser dengan tampilan tertentu (Informatika, UPT Sumber Daya 2017). Setiap *tag* html mempunyai atribut yang berbeda-beda di dalamnya, atribut ini digunakan untuk melakukan pengaturan tambahan terhadap *tag* yang memilikinya. Ada tiga *tag* utama yang membangun sebuah halaman html, yaitu *tag* html (`<html> </html>`), *tag* head (`<head> </head>`) dan *tag* body (`<body> </body>`).

Tag body adalah *tag* yang mengapit semua informasi yang ingin ditampilkan pada sebuah halaman *website*. Situs berita pada umumnya menempatkan isi berita di antara *tag* p (*tag* untuk menampilkan paragraf) yang ada di dalam *tag* body. Untuk mengambil berita tersebut, maka harus dilakukan proses ekstraksi untuk memisahkan konten berita dari *tag* html yang mengapitnya, proses ekstraksi ini lebih dikenal dengan istilah *Web Scraping*.

2.3 *Web Scraping*

Web scraping menggambarkan penggunaan sebuah program untuk mengekstraksi data dari file HTML yang ada di sebuah situs tertentu di internet. Pada penelitian yang dilakukan Haddaway pada tahun 2015 menjelaskan bahwa *web scraping* adalah istilah yang digunakan untuk menggambarkan proses ekstraksi

data pada *file* elektronik menggunakan sebuah program komputer (Haddaway 2015). Penelitian lain yang dilakukan oleh Utomo pada tahun 2013 menjelaskan bahwa teknik *scraping* dapat dilakukan dengan berbagai cara, di antaranya menggunakan analisa DOM-HTML dan dengan menggunakan teknik *website* ekspresi reguler (*Regex*), kedua teknik ini mempunyai keunggulan tersendiri dan menghasilkan hasil yang tidak jauh berbeda (Utomo 2013).

Web scraping mempunyai tahapan-tahapan di dalamnya, Maria Rosario B, dkk pada tahun 2017 menjelaskan bahwa secara umum *web scraping* memiliki beberapa langkah yaitu membuat *scraping template*, melakukan eksplorasi navigasi situs, mengotomatisasi navigasi dan mengekstrak informasi serta menyimpannya (Maria Rosario B, Yovi Pratama, and Fachruddin 2014).

1. Membuat *template scraping*

Proses ini melakukan observasi terhadap dokumen HTML situs dijadikan objek *scraping*. Cara yang digunakan adalah dengan menganalisis *tag* HTML yang mengapit informasi yang akan di ambil.

2. Eksplorasi navigasi situs

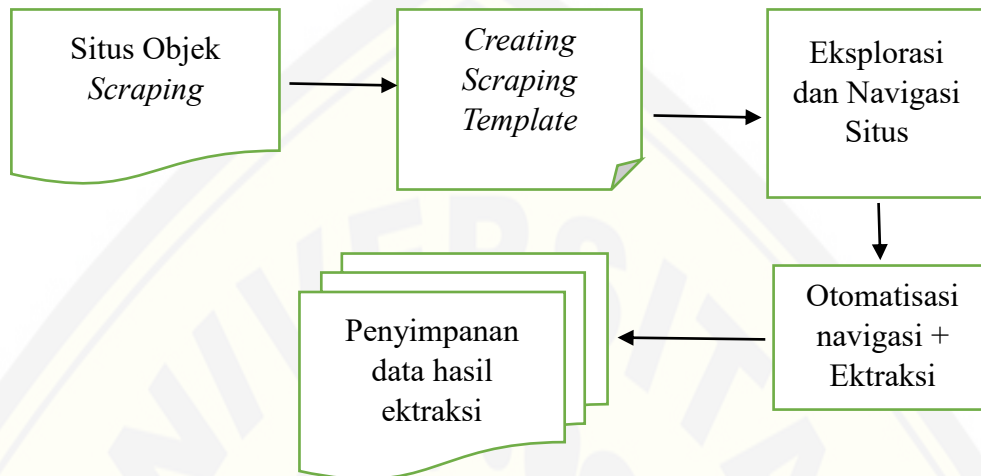
Proses ini dilakukan dengan menelusuri navigasi pada situs objek *scraping* untuk ditirukan oleh *web scraper* yang akan dibangun.

3. Mengotomatisasi navigasi

Berdasarkan hasil dari langkah satu dan dua, aplikasi dibangun untuk mengotomatisasi pengambilan informasi pada situs objek *scraping*.

4. Ekstraksi data dan menyimpan histori

Informasi yang diekstrak dari langkah tiga disimpan pada tabel-tabel di dalam *basis data*. Cara kerja teknik *scraping* dapat dilihat pada gambar 2.1.



Gambar 2. 1 Langkah-langkah *Web Scraping*

Gambar 2.1 menggambarkan langkah-langkah melakukan *web scraping*, dimulai dengan menentukan situs yang akan dijadikan objek *scraping*, kemudian menganalisis struktur halaman situs tersebut untuk membuat *template*. Langkah selanjutnya adalah melakukan eksplorasi dan navigasi, hasil dari proses ini kemudian di otomatisasi untuk melakukan ekstraksi. Langkah terakhir adalah menyimpan hasil *scraping* yang telah dilakukan. Proses ekstraksi informasi sebelum disimpan dapat dilakukan dengan memanfaatkan metode ekspresi reguler untuk menemukan informasi yang diinginkan dengan pola sesuai *template* yang sudah dibuat.

2.4 Ekspresi Reguler

Selain menggunakan metode DOM, *web sraping* juga bisa dilakukan dengan menggunakan metode *regular expression (Regex)*. *Regular expression* atau yang sering disebut sebagai *Regex* adalah sebuah formula untuk pencarian pola suatu kalimat/string (Muliantara 2009). Muliantara tersebut juga menjelaskan beberapa pola umum yang digunakan pada ekspresi reguler, yang bisa dilihat pada tabel 2.2.

Tabel 2. 1 Pola Ekspresi Reguler

Pola	Penjelasan
[]	Ekspresi kurung tegak, digunakan untuk mencari salah satu karakter tertentu yang berada di dalam kurung, misal : <i>pattern</i> “a[efg]i” adalah <i>pattern</i> yang

Pola	Penjelasan
	digunakan untuk mencari <i>string</i> / kata dengan karakter e / f / g di antara a dan i, <i>pattern</i> tersebut akan cocok pada <i>string</i> “aei”, “afi”, atau “agi”.
[^]	Mencari sebuah karakter yang tidak ada di dalam kurung, berlawanan dengan pola “[]”. misal : <i>pattern</i> “[^def]”, <i>pattern</i> tersebut akan cocok pada semua karakter, kecuali karakter d, e, dan f.
?	Pola ini digunakan untuk mencari nol atau satu karakter sebelumnya, misal : <i>pattern</i> “ab?c”, <i>pattern</i> tersebut akan cocok pada <i>string</i> “ac” dan “abc”.
+	Pola ini digunakan untuk mencari 1 atau lebih karakter sebelumnya, misal : <i>pattern</i> “ab+c” akan cocok pada <i>string</i> “abc, abbc, abbbbc, dst.”
*	Pola ini digunakan untuk mencari 0 atau lebih karakter sebelumnya, misal : <i>pattern</i> “ab*c” akan cocok pada <i>string</i> “ac, abc, abbc, abbbbc, dst.”
.	Mencari karakter apa saja kecuali newline (\n)
!	Jika diletakkan di depan <i>pattern</i> maka berarti “bukan”, contoh : <i>pattern</i> “!a.u” maka akan cocok dengan <i>string</i> apa saja kecuali “aku, anu, alu, agu, dan seterusnya”.
^	Jika diletakkan di depan <i>pattern</i> maka akan cocok dengan awal sebuah <i>string</i>
()	<i>Grouping</i> , digunakan untuk mengelompokkan karakter-karakter menjadi <i>single unit</i> .
\	<i>Escape Character</i> mengembalikan fungsi <i>meta character</i> menjadi karakter biasa.

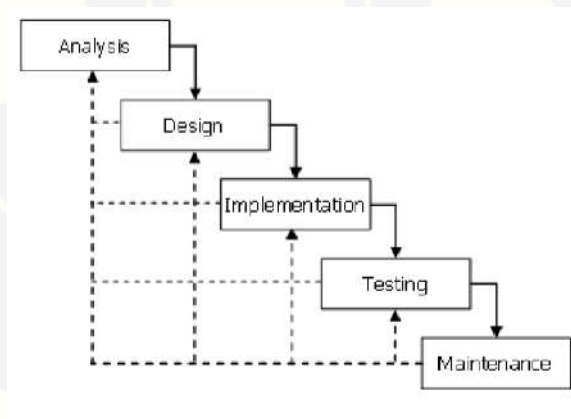
Pola-pola pada tabel 2.2 dapat digunakan untuk mencari dan mengambil informasi dari dalam struktur html sebuah halaman *website*. Pola *regex* yang dibentuk diimplementasikan pada aplikasi yang dikembangkan.

2.5 Software Development Life Cycle (SDLC) Waterfall

Pembangunan aplikasi dilakukan dengan beberapa tahapan yaitu analisis kebutuhan, desain untuk perancangan aplikasi dan implementasi sesuai dengan analisis dan desain yang telah dirancang. Dalam pembangunan aplikasi dibutuhkan sebuah pendekatan untuk pengembangan perangkat lunak berbasis komputer yang dikenal sebagai *System Development Life Cycle* (SDLC). SDLC merupakan suatu

urutan dari beberapa proses secara bertahap di dalam merancang dan mengembangkan aplikasi / aplikasi yang dikenal dengan nama *Information System Development* atau juga *Application Development*.

Ada beberapa pendekatan SDLC yang digunakan untuk pengembangan aplikasi, di antaranya adalah *waterfall*. Penelitian yang dilakukan oleh Youssef Bassil pada tahun 2012 menjelaskan bahwa *waterfall* adalah sebuah model yang memiliki tahapan yang harus dilakukan secara berurutan agar proses pengembangan aplikasi dapat berhasil (Youssef Bassil 2012). Tahapan pada SDLC ini dapat dilihat pada gambar 2.2.



Gambar 2. 2 Model *Waterfall* (Youssef Bassil 2012)

Gambar 2.2 menggambarkan tahapan-tahapan pada proses pengembangan aplikasi dengan pendekatan *waterfall*. Tahapan pertama adalah analisis, pada tahap ini dilakukan analisis kebutuhan aplikasi, desain adalah tahapan kedua setelah tahap analisis selesai dilakukan, dilanjutkan dengan implementasi kode, setelah tahap implementasi kode selesai, tahap selanjutnya adalah *testing* yaitu tahap pengujian terhadap aplikasi yang telah dikembangkan, setelah tahap *testing* selesai dilanjutkan dengan tahapan *maintenance* sebagai tahap akhir pengembangan.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tentang gambaran tahapan yang dilakukan untuk menganalisis data untuk menjawab perumusan masalah sehingga dapat mencapai tujuan sebenarnya dari penelitian. Pada metodologi penelitian akan dijelaskan tentang tahapan dari penelitian

3.1 Jenis Penelitian

Penelitian ini menggunakan jenis penelitian pengembangan. Penelitian ini bertujuan untuk menemukan pola *Regex* untuk ekstraksi informasi dari tiga situs berita dan menemukan perbedaan pola pada masing-masing situs. Untuk mengimplementasikan pola *Regex* maka akan dikembangkan aplikasi yang berfungsi untuk mengambil data secara otomatis dari situs berita www.detik.com, www.tribunnews.com dan www.liputan6.com berdasarkan pola *Regex* masing-masing situs. Data yang berhasil diambil akan dijadikan *repository* berita yang dapat diakses melalui API yang disediakan oleh aplikasi yang akan dikembangkan.

3.2 Waktu Penelitian

Waktu penelitian dilakukan selama 7 bulan dimulai pada bulan Desember 2017 sampai dengan bulan Juli 2018.

3.3 Gambaran Aplikasi

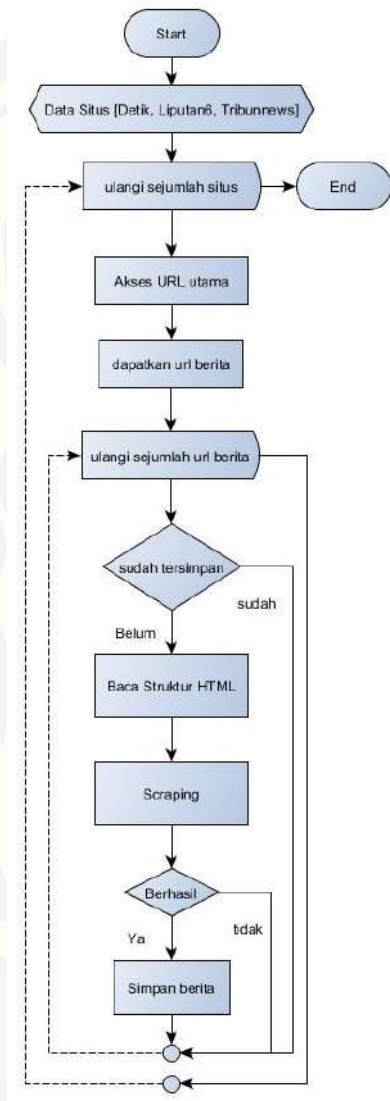
Berdasarkan fungsionalitas aplikasi pada batasan masalah, maka aplikasi akan dikembangkan dengan mengimplementasikan teknik *scraping website* menggunakan metode *regular expression (Regex)* untuk memisahkan konten dari struktur HTML halaman situs yang dituju. Situs yang akan dijadikan objek *scraping* adalah tiga situs berita Indonesia yang memiliki *ranking* tiga teratas di www.alexacom (www.tribunnews.com, www.detik.com, dan www.liputan6.com). Informasi yang akan di ambil dari ketiga *website* tersebut adalah judul berita, penulis/sumber berita, tanggal publikasi berita, alamat berita, isi berita dan tautan berita berupa video. Aplikasi ini akan berjalan dengan menyimulasikan akses ke

website tersebut seperti aplikasi *browser* atau sering disebut *exploration site navigation*. Aplikasi akan mengunjungi satu persatu halaman yang memuat informasi yang dibutuhkan.

Setelah proses *exploration site navigation* selesai, data yang didapat akan di simpan ke dalam basis data. Tujuan utama aplikasi ini adalah untuk mengumpulkan data berita di Indonesia sehingga dapat digunakan sebagai sumber data analisis untuk penelitian-penelitian lain. Aplikasi akan dibangun menjadi dua sub aplikasi yaitu *scraper* dan *monitoring*.

3.3.1. Scraper

Sub aplikasi ini berfungsi untuk mengekstraksi data-data yang dibutuhkan dari situs berita. Tugas utama dari sub aplikasi ini adalah memisahkan konten yang dibutuhkan dari struktur HTML yang dimiliki oleh situs berita. Secara garis besar alur kerja *scraper* dapat dilihat pada gambar 3.1.



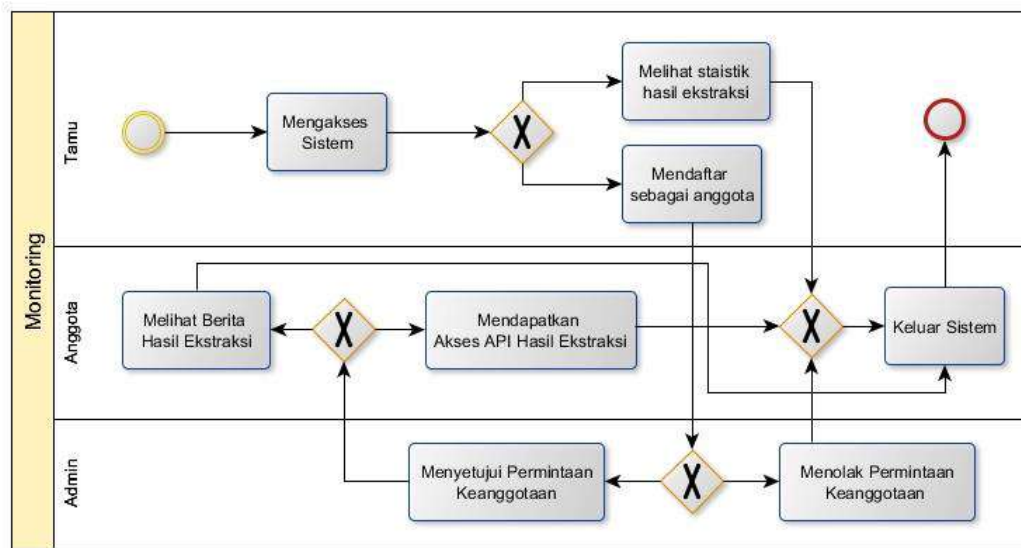
Gambar 3. 1 Alur Kerja Aplikasi

Gambar 3.1 menggambarkan alur kerja dari *scraper* dimulai dengan mengakses alamat halaman indeks dari masing-masing situs, kemudian mengekstrak alamat dari masing-masing berita yang ada pada halaman indeks. Alamat yang berhasil didapatkan kemudian diperiksa, jika alamat itu masih belum pernah di ekstrak, maka alamat tersebut akan dikunjungi untuk di ekstrak informasi di dalamnya, setelah itu di simpan ke dalam basis data, jika sudah tersimpan maka dilanjutkan ke alamat berita selanjutnya.

3.3.2. Monitoring

Sub aplikasi ini berfungsi sebagai media *monitoring* hasil ekstraksi yang

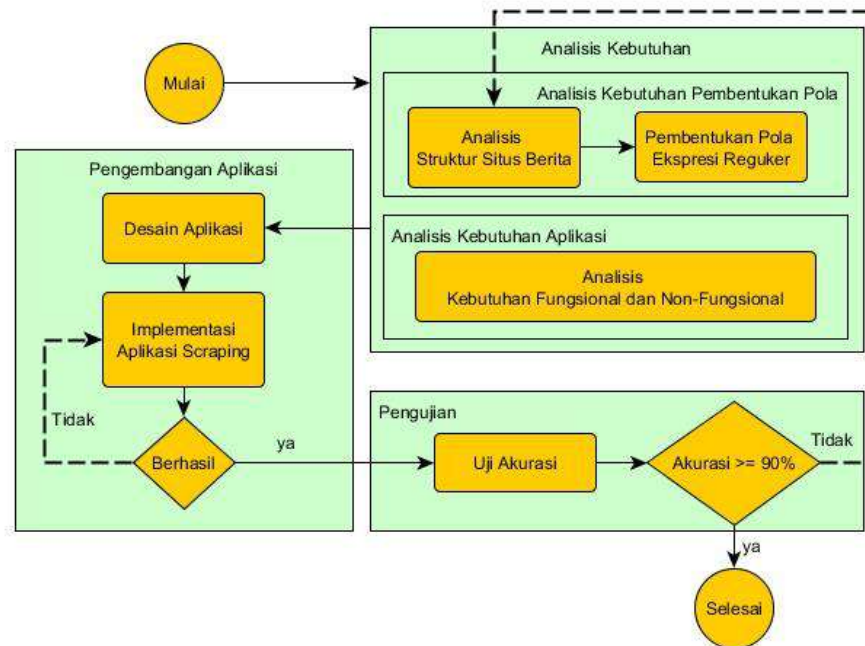
telah dilakukan *scraper*. Tugas utama sub aplikasi ini adalah menampilkan data hasil proses *scraping* agar mudah dipahami oleh pengguna, membantu admin dalam pengelolaan anggota dan akses *Application Programming Interface* (API) serta mempermudah admin untuk mengelola pola dari ekspresi reguler yang digunakan oleh *scraper*. Secara umum proses bisnis pada sub aplikasi *monitoring* dapat dilihat pada gambar 3.2



Gambar 3. 2 Proses bisnis sub aplikasi *monitoring*

3.4 Tahapan Penelitian

Penelitian ini dilakukan dalam beberapa tahap, tahapan dalam penelitian ini dapat dilihat pada gambar 3.3.



Gambar 3. 3 Tahapan Penelitian

Gambar 3.3 menggambarkan tahapan penelitian yang akan dilakukan peneliti. Langkah awal yang dilakukan adalah melakukan analisis kebutuhan, yaitu analisis kebutuhan pembentukan pola dan analisis kebutuhan aplikasi. Analisis kebutuhan pembentukan pola dilakukan melalui dua tahapan, yaitu analisis struktur halaman situs dan pembentukan pola *regex*. Analisis kebutuhan aplikasi dilakukan secara paralel dengan analisis kebutuhan pembentukan pola *regex*, kemudian dilanjutkan dengan desain dan implementasi aplikasi *scraping* berdasarkan hasil analisis kebutuhan. Setelah aplikasi berhasil dikembangkan maka dilakukan pengujian dengan melakukan uji akurasi, jika hasil uji akurasi kurang dari 90%, maka dilakukan proses analisis struktur halaman ulang hingga hasil uji akurasi $\geq 90\%$.

3.4.1. Analisis Kebutuhan Pembentukan Pola *Regex*

Pada tahapan ini dilakukan dua tahapan penelitian yang dilakukan secara paralel, yaitu tahap analisis struktur situs berita dan pembentukan pola *regex*.

1. Analisis Struktur Situs Berita

Tahapan ini dilakukan dengan melakukan analisis terhadap struktur HTML yang digunakan situs berita. Pada umumnya, situs berita memiliki dua halaman

utama, yaitu halaman indeks dan halaman detail berita. Halaman indeks adalah halaman yang berisi daftar berita terbaru yang di publikasi setiap harinya. Struktur halaman ini di analisis untuk mengekstraksi *tautan* berita yang di publikasi. Contoh struktur HTML halaman indeks berita dapat dilihat pada gambar 3.4.

```

<article>
  <a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
  Berita" href="https://m.detik.com/news/detiktv/d-3789245/kejam-bom-bunuh-diri-di-
  afghanistan-tewaskan-puluhan-orang" class="list">...</a>
</article>
<article>
  <a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
  Berita" href="https://m.detik.com/news/berita-jawa-tengah/d-3789241/ada-33-kasus-
  bunuh-diri-di-gunungkidul-gantung-diri-yang-terbanyak" class="list">...</a>
</article>
<article>
  <a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
  Berita" href="https://m.detik.com/news/berita-jawa-tengah/d-3789214/congkel-jok-
  motor-milik-polwan-dua-residivis-dibekuk-polisi" class="list">...</a>
</article>

```

Gambar 3. 4 Struktur halaman indeks berita Detik.com

Gambar 3.4 adalah struktur halaman indeks dari situs www.detik.com. Halaman ini berisi kumpulan judul dan *tautan* berita terbaru yang di publikasikan oleh detik. Halaman ini di analisis untuk mendapatkan pola peletakan *tautan* dari masing-masing berita yang ditampilkan.

Halaman yang di analisis selanjutnya adalah halaman detail berita, yaitu semua halaman yang dituju oleh *tautan* pada halaman indeks. Halaman ini adalah halaman yang berisi konten berita, tanggal publikasi, dan penulis berita. Analisis pada halaman ini dilakukan untuk mendapatkan pola HTML yang digunakan untuk menampilkan berita yang akan di ekstraksi. Salah satu contoh struktur halaman berita dapat dilihat pada gambar 3.5.

```

<article>
  <!--:JUDUL ARTICLE-->
  <div class="detail_area">
    <div class="date">Kamis 28 Desember 2017, 18:33 WIB</div>
    <div class="sub"></div>
    <h1 class="jdl">Ada 34 Kasus Bunuh Diri di Gunungkidul, Gantung Diri yang
    Terbanyak</h1>
    <div class="author">...</div>
    <!-- <div class="author">
    <strong>Usman Hadi</strong> - detikNews
    </div> -->
    <div class="clearfix">...</div>
  </div>
  <!--E:JUDUL ARTICLE-->
  <div class="pic">...</div>
  <!--S:ISI ARTICLE-->
  <div class="text_detail detail_area" id="detikdetailtext">
    <strong>Gunungkidul</strong>
    "
    "
    Tercatat ada 34 kasus bunuh diri di Gunungkidul, Daerah Istimewa Yogyakarta
    (DIY) selama tahun 2017. Kasus gantung diri menjadi yang terbanyak."
    <br>
    <br>
    ""Tahun ini ada 34 kasus bunuh diri, meningkat dibandingkan tahun 2016 lalu
    yakni 33 kasus," kata Kapolres Gunungkidul, AKBP Ahmad Fuadi kepada wartawan
    di Mapolres, Kamis (28/12/2017).""
    <br>
    <br>
    "Mayoritas kasus bunuh diri di Gunungkidul, kata Ahmad, didominasi kasus
    gantung diri dengan 29 kasus. Kemudian percobaan bunuh diri 4 kasus dan orang
    bunuh diri dengan masuk ke dalam sumur 1 kasus."
  </div>

```

Gambar 3. 5 Struktur halaman berita Detik.com

Gambar 3.5 adalah contoh dari struktur halaman salah satu berita yang di publikasikan di www.detik.com. Halaman ini berisi informasi berupa judul berita, penulis/sumber berita, tanggal publikasi berita, alamat berita, dan isi berita. Analisis pada halaman ini diperlukan untuk mendapatkan pola peletakan masing-masing informasi yang akan di ambil. Pada tahapan ini dilakukan observasi pada struktur html pada masing-masing situs untuk menemukan struktur penulisan dari enam konten yang akan diambil.

a. Penulisan *tautan* berita.

Observasi ini dilakukan pada struktur html pada halaman indeks situs. Observasi yang dilakukan akan menghasilkan pola *Regex* yang berfungsi untuk mengambil *tautan* berita dari halaman indeks situs.

b. Penulisan judul berita.

Observasi ini dilakukan pada struktur html pada halaman berita situs. Observasi yang dilakukan akan menghasilkan pola *Regex* yang berfungsi untuk mengambil judul berita.

c. Penulisan isi berita.

Observasi ini dilakukan pada struktur html pada halaman berita situs. Observasi yang dilakukan akan menghasilkan pola *Regex* yang berfungsi untuk mengambil isi berita.

d. Penulisan penulis berita.

Observasi ini dilakukan pada struktur html pada halaman berita situs. Observasi yang dilakukan akan menghasilkan pola *Regex* yang berfungsi untuk mengambil penulis/sumber berita.

e. Penulisan tanggal publikasi berita.

Observasi ini dilakukan pada struktur html pada halaman berita situs. Observasi yang dilakukan akan menghasilkan pola *Regex* yang berfungsi untuk mengambil tanggal publikasi berita.

f. Penulisan *tautan* berita berupa video.

Observasi ini dilakukan pada struktur html pada halaman berita situs. Observasi yang dilakukan akan menghasilkan pola *Regex* yang berfungsi untuk mengambil *tautan* berita berupa video.

2. Pembentukan Pola Ekspresi Reguler

Tahapan ini adalah tahapan pembentukan pola ekspresi reguler berdasarkan hasil dari tahap analisis struktur halaman situs berita. Secara garis besar akan dibentuk pola ekspresi reguler untuk mengekstrak enam konten yang berbeda-beda. Contoh struktur penulisan tautan berita situs detik.com dapat dilihat pada gambar 3.6

```
><a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
Berita" href="https://m.detik.com/news/detiktv/d-3789245/kejam-bom-bunuh-diri-di-
afghanistan-tewaskan-puluhan-orang" class="list">...</a>
```

Gambar 3. 6 Struktur Penulisan Tautan Berita Detik.com

Gambar 3.6 adalah contoh struktur penulisan tautan berita pada halaman indeks situs www.detik.com. Hasil observasi menunjukkan bahwa *tautan* berita terletak pada atribut “*href*” yang dimiliki oleh *tag* <a>. Berdasarkan observasi struktur HTML pada gambar 3.5, untuk mengekstraksi *tautan* berita digunakan pola :

```
<a data-category="Subkanal detikNews" data-action="Indeks" data-label="List Berita"
href="(.*?)" class="list">
```

Pola *Regex* untuk proses ekstraksi enam konten berita akan ditentukan menggunakan cara yang sama seperti menentukan pola penulisan *tautan* yang sudah dicontohkan sebelumnya. Enam konten berita tersebut adalah judul berita, penulis/sumber berita, tanggal publikasi berita, alamat berita, isi berita dan tautan dari berita berupa video.

3.4.2. Analisa Kebutuhan Aplikasi

Analisa kebutuhan aplikasi dilakukan berdasarkan fungsional aplikasi yang akan dibangun sesuai dengan batasan masalah. Tahap ini akan menghasilkan kebutuhan fungsional dan non-fungsional apa saja yang dibutuhkan untuk memenuhi fungsionalitas aplikasi sesuai dengan batasan masalah. Pada tahap ini juga akan di analisis aktor siapa saja yang akan berinteraksi dengan aplikasi dan kebutuhan fitur apa saja yang menunjang pengelolaan aktor tersebut.

3.4.3. Desain Aplikasi

Hasil dari tahapan analisis kebutuhan akan digabungkan dengan hasil tahapan penelitian 3.4.1 dan 3.4.2 akan digunakan untuk tahapan perancangan

aplikasi. Perancangan aplikasi akan dilakukan menggunakan *Unified Modeling Language* (UML) yang terdiri atas :

- 1) *Bussiness Process*: *Bussiness Process* digunakan untuk menggambarkan proses dari sebuah aplikasi yang meliputi *input*, *output*, dan *goal* yang merupakan tujuan dari sebuah aplikasi yang dibangun.
- 2) *Usecase*: *Usecase* digunakan untuk merepresentasikan interaksi antara pengguna dan aplikasi dalam UML. *Usecase* menggambarkan interaksi antara *user* dengan aplikasi, dan dapat menggambarkan hak akses *user*.
- 3) *Skenario*: *Skenario* digunakan untuk menjelaskan alur proses untuk setiap fungsionalitas yang ada pada *usecase* diagram.
- 4) *Activity Diagram*: *Activity Diagram* digunakan untuk menggambarkan berbagai alur aktivitas dalam aplikasi yang sedang dirancang, bagaimana masing-masing alur berawal, *decision* yang mungkin terjadi, dan bagaimana mereka berakhir.
- 5) *Sequence Diagram* : *Sequence Diagram* (diagram urutan) digunakan untuk memperlihatkan atau menampilkan interaksi-interaksi antar objek di dalam aplikasi yang disusun pada sebuah urutan atau rangkaian waktu. Interaksi antar objek tersebut termasuk pengguna, *display*, dan sebagainya berupa pesan atau *message*.
- 6) *Class Diagram*: *Class Diagram* digunakan untuk menggambarkan struktur dan deskripsi class, package dan objek beserta hubungan satu sama lain seperti pewarisan, asosiasi, dan lain-lain.
- 7) *Entity Relationship Diagram*: *Entity Relationship Diagram* digunakan untuk menggambarkan struktur basis data yang akan dibangun pada aplikasi.

3.4.4. Implementasi *Web Scraping*

Proses yang selanjutnya dilakukan adalah proses implementasi kode (*coding*). Tahapan ini akan mengimplementasikan perancangan yang sudah dibuat pada tahapan sebelumnya ke dalam bahasa *website*. Bahasa *website* yang digunakan adalah *c#* yang akan diintegrasikan dengan DBMS *postgresql* sebagai media penyimpanan datanya. Tahapan ini juga mencakup dua tahapan *scraping* yang paling penting yaitu:

- 1) Mengotomatisasi eksplorasi navigasi yang telah dirancang sebelumnya.

Berdasarkan hasil dari perancangan sebelumnya, penulisan kode *website* dilakukan untuk mengotomatisasi pengambilan informasi pada situs objek *scraping*

- 2) Menyimpan data hasil ekstraksi ke dalam *basis data* Informasi yang di dapat dari langkah tiga disimpan pada tabel-tabel di dalam basis data.

Pada tahap pengembangan aplikasi ini, komputer yang digunakan oleh peneliti mempunyai spesifikasi sebagai berikut:

- 1) Prosesor *core-i5*
- 2) RAM (*memory*) 4GB
- 3) Sistem Operasi *Windows 10*

Aplikasi *scrapper* yang dikembangkan akan dijalankan setiap lima menit pada server dengan spesifikasi minimal sebagai berikut:

- 1) Prosesor *core-i5*
- 2) RAM (*memory*) 4GB
- 3) Sistem Operasi *Windows 7*

3.4.5. Uji Akurasi

Uji akurasi merupakan suatu cara untuk menghitung seberapa baik suatu aplikasi dalam menemukan dokumen yang sesuai dengan kebutuhan pengguna. Penilaian yang umum dilakukan adalah dengan menggunakan metode perbandingan. Pada penelitian yang dilakukan oleh Rila Mandala pada tahun 2006 (Mandala, 2006), digunakan dua metode perbandingan yaitu *precision* dan *recall* untuk mengukur akurasi dari efektivitas metode *machine-learning* pada *search-engine*.

Pada penelitian yang dilakukan oleh Rila Mandala pada tahun 2006 dijelaskan bahwa *precision* adalah perbandingan jumlah dokumen relevan yang diambil dengan jumlah seluruh dokumen yang diambil oleh aplikasi (Mandala 2006). *Recall* adalah perbandingan antara jumlah dokumen relevan yang diambil dengan jumlah dokumen relevan yang berada di koleksi dokumen (*basis data*). Untuk menghitung nilai *precision* digunakan persamaan 1) dan untuk menghitung nilai *recall* digunakan persamaan 2).

$$Precision = \frac{\text{Relevan document retrieved}}{\text{Retrieved document}} \dots\dots\dots 1)$$

$$Recall = \frac{\text{Relevan document retrieved}}{\text{Relevant document}} \dots\dots\dots 2)$$

Pada tahap uji akurasi ini akan dilakukan pengujian dengan menggunakan 2 metode perbandingan tersebut, yaitu metode *precision* untuk mengetahui sejauh mana performa aplikasi yang dibangun dilihat dari jumlah data valid yang diambil jika dibandingkan dengan jumlah data yang di ambil. *Recall* digunakan untuk mengetahui tingkat akurasi aplikasi jika dilihat dari jumlah data valid yang diambil dibandingkan dengan jumlah data yang seharusnya diambil.

Hasil dari perhitungan *precision* dan *recall* akan digunakan untuk menghitung *F-Measure*. *F-Measure* adalah kombinasi rata-rata *harmonic* dari *presicion* dan *recall* yang berbanding lurus dengan nilai keduanya. Nilai *recall* dan *precision* pada suatu keadaan dapat memiliki bobot yang berbeda. Rumus mencari *F-Measure* dapat ditulis sebagai Persamaan 3) berikut ini (Ling, N. Kencana, and Oka 2014).

$$F_{measure} = \frac{2*precision*recall}{precision+recall} \dots\dots\dots 3)$$

BAB 4 PENGEMBANGAN APLIKASI

Bab ini akan membahas tentang pengembangan aplikasi ekstraksi web (*web scraping*) dengan mengimplementasikan metode ekspresi reguler (*regular expression*). Tahap pengembangan dilaksanakan berdasarkan tahapan penelitian.

4.1. Analisis Struktur Situs Berita

Tahapan analisis struktur situs berita adalah tahapan awal yang dilakukan peneliti. Tahapan ini dengan melakukan observasi terhadap struktur html masing-masing situs. Observasi ini dilakukan guna mencari struktur penulisan enam konten yang ingin di ekstrak, yaitu *tautan*, judul, penulis, tanggal publikasi, isi berita dan tautan berita berupa video.

4.1.1. Observasi Penulisan *Tautan* Berita

Observasi ini dilakukan untuk menganalisis struktur penulisan *tautan* berita pada halaman indeks masing-masing situs. Pada tahap ini peneliti menggunakan struktur halaman dengan *user-agent* berupa *mobile* sesuai dengan batasan masalah.

1. Detik

Detik memiliki dua jenis berita dengan halaman indeks yang berbeda, yaitu halaman indeks untuk berita narasi dan halaman indeks untuk berita video. Kedua halaman tersebut memiliki pola struktur html yang sama, jadi hanya perlu dilakukan satu kali analisis pada salah satu halaman. Pada penelitian ini, penulis melakukan analisis terhadap halaman indeks dari berita narasi.

```

▼<article>
  ▶<a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
  Berita" href="https://m.detik.com/news/detiktv/d-3789245/kejam-bom-bunuh-diri-di-
  afghanistan-tewaskan-puluhan-orang" class="list">...</a>
</article>
▼<article>
  ▶<a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
  Berita" href="https://m.detik.com/news/berita-jawa-tengah/d-3789241/ada-33-kasus-
  bunuh-diri-di-gunungkidul-gantung-diri-yang-terbanyak" class="list">...</a>
</article>
▼<article>
  ▶<a data-category="Subkanal detikNews" data-action="Indeks" data-label="List
  Berita" href="https://m.detik.com/news/berita-jawa-tengah/d-3789214/congkel-jok-
  motor-milik-polwan-dua-residivis-dibekuk-polisi" class="list">...</a>
</article>

```

Gambar 4. 1 Struktur halaman indeks www.detik.com

Gambar 4.1 adalah struktur penulisan tautan berita pada halaman indeks situs www.m.detik.com. Semua judul berita ditulis menggunakan *tag* `<a>` yang

mengandung tautan menuju halaman berita tertentu memiliki jumlah atribut tambahan dan nilai atribut yang sama. Secara umum penulisan tautan berita pada halaman indeks www.detik.com dengan *user-agent mobile* adalah `<a data-category="Subkanal detikNews" data-cation="Indeks" data-label="List Berita" href="Tautan Menuju Halaman Berita" class="list">Judul Berita Yang Dituju`. Pola penulisan seperti ini hanya digunakan untuk menuliskan tautan pada judul berita pada halaman indeks, sehingga dapat dijadikan acuan untuk melakukan ekstraksi tautan berita.

2. Liputan 6

```

▼<div class="article-snippet js-article-snippet" data-id="3571134" data-type="TextTypeArticle" data-channel="news" data-label data-template-name="article-snippet" data-template-var="container" data-component-name="mobile:article-snippet">
  ▶<div class="article-snippet__image js-article-snippet-image">...</div>
  ▶<div class="article-snippet__content js-article-snippet-content" data-template-var="content">
    ▼<div class="article-snippet__info">
      ▶<div class="article-snippet__channel-wrapper" data-template-var="channel">...</div>
      ▶<a href="https://m.liputan6.com/news/read/3571134/trauma-dengan-insiden-danau-toba-213-penumpang-kapal-diturunkan-paksa" class="article-snippet__title-link" title="Trauma dengan Insiden Danau Toba, 213 Penumpang Kapal Diturunkan Paksa" target="_self" data-sponsored data-template-var="title-link">...</a>
    </div>
    ▶<div class="article-snippet__footer" data-template-var="footer">...</div>
  </div>
  </div>
▼<div class="article-snippet js-article-snippet" data-id="3571271" data-type="TextTypeArticle" data-channel="news" data-label data-template-name="article-snippet" data-template-var="container" data-component-name="mobile:article-snippet">
  ▶<div class="article-snippet__image js-article-snippet-image">...</div>
  ▶<div class="article-snippet__content js-article-snippet-content" data-template-var="content">
    ▼<div class="article-snippet__info">
      ▶<div class="article-snippet__channel-wrapper" data-template-var="channel">...</div>
      ▶<a href="https://m.liputan6.com/news/read/3571271/bahaya-balon-udara-terhadap-jaringan-listrik" class="article-snippet__title-link" title="Bahaya Balon Udara Terhadap Jaringan Listrik" target="_self" data-sponsored data-template-var="title-link">...</a>
    </div>
    ▶<div class="article-snippet__footer" data-template-var="footer">...</div>
  </div>
  </div>
▼<div class="article-snippet js-article-snippet" data-id="3571176" data-type="TextTypeArticle" data-channel="news" data-label data-template-name="article-snippet" data-template-var="container" data-component-name="mobile:article-snippet">
  ▶<div class="article-snippet__image js-article-snippet-image">...</div>
  ▶<div class="article-snippet__content js-article-snippet-content" data-template-var="content">
    ▼<div class="article-snippet__info">
      ▶<div class="article-snippet__channel-wrapper" data-template-var="channel">...</div>
      ▶<a href="https://m.liputan6.com/news/read/3571176/tjahjo-terdaftar-di-pilkada-jateng-meski-ber-ktp-dki-ini-kata-kpu" class="article-snippet__title-link" title="Tjahjo Terdaftar di Pilkada Jateng Meski Ber-KTP DKI, Ini Kata KPU" target="_self" data-sponsored data-template-var="title-link">...</a>
    </div>
    ▶<div class="article-snippet__footer" data-template-var="footer">...</div>
  </div>
  </div>
</div>

```

Gambar 4. 2 Struktur halaman indeks www.liputan6.com

Gambar 4.2 adalah struktur penulisan tautan berita pada halaman indeks situs www.liputan6.com. Situs berita www.liputan6.com menambahkan atribut tambahan berupa *class*, *title*, *target* dan *data-sponsored*, *data-template-var* pada tag `<a>` yang digunakan. Secara umum penulisan tautan berita pada halaman indeks www.liputan6.com dengan *user-agent mobile* adalah `</div>Judul Berita Yang Dituju`. Pola penulisan

seperti ini hanya digunakan untuk menuliskan tautan pada judul berita pada halaman indeks, sehingga dapat dijadikan acuan untuk melakukan ekstraksi tautan berita.

3. Tribunnews

```

▼<li sort="3">
  ▶<div class="fr btsquare pos_rel">...</div>
  ▼<div class="m-btsquare">
    ▶<h4 class="fbo2 f13 red">...</h4>
    ▼<h3 class="blue">
      ▶<a href="http://m.tribunnews.com/regional/2018/06/27/pantau-hasil-perhitungan-suara-cawabup-bandung-barat-hengky-kurniawan-curhat-di-medsos" title="Pantau Hasil Perhitungan Suara, Cawabup Bandung Barat Hengky Kurniawan Curhat Di Medsos" class="fbo2 f16 ln22 txt-oev-3">...</a>
      </h3>
    ▶<div class="grey pt5 dip">...</div>
  </div>
  <div class="c12"></div>
</li>
▼<li sort="4">
  ▶<div class="fr btsquare pos_rel">...</div>
  ▼<div class="m-btsquare">
    ▼<h3 class="blue">
      ▶<a href="http://m.tribunnews.com/nasional/2018/06/27/soal-penggalangan-dana-psi-sekarang-gerindra-mengekor-langkah-kami" title="Soal Penggalangan Dana, PSI: Sekarang Gerindra Mengekor Langkah Kami" class="fbo2 f16 ln22 txt-oev-3">...</a>
      </h3>
    ▶<div class="grey pt5 dip">...</div>
  </div>
  <div class="c12"></div>
</li>
▼<li sort="5">
  ▶<div class="fr btsquare pos_rel">...</div>
  ▼<div class="m-btsquare">
    ▼<h3 class="blue">
      <a href="http://m.tribunnews.com/regional/2018/06/27/kisah-tki-asal-bali-meninggal-di-turki-ini-firasat-sang-bunda" title="Kisah TKI Asal Bali Meninggal di Turki, Ini Firasat Sang Bunda" class="fbo2 f16 ln22 txt-oev-3">
        Kisah TKI Asal Bali Meninggal di Turki, Ini Firasat Sang Bunda
      </a>
    </h3>
    ▶<div class="grey pt5 dip">...</div>
  </div>
  <div class="c12"></div>
</li>

```

Gambar 4. 3 Struktur halaman indeks www.tribunnews.com

Gambar 4.3 adalah struktur penulisan tautan berita pada halaman indeks situs www.tribunnews.com. Situs berita www.tribunnews.com menambahkan atribut tambahan berupa *title*, dan *class* pada `<a>` yang memuat tautan menuju halaman berita tertentu. Semua *tag* `<a>` yang mengandung tautan menuju halaman berita tertentu memiliki jumlah atribut tambahan dan nilai atribut yang sama yaitu atribut *class* dengan *value* "fbo2 f16 ln22 txt-oev-3" dan *title* yang memiliki *value* berbeda-beda pada setiap berita. Secara umum penulisan tautan berita pada halaman indeks www.tribunnews.com dengan *user-agent mobile* adalah `<h3 class="blue">Judul Berita Yang Dituju</h3>`.

4.1.2. Observasi Penulisan Judul Berita

1. Detik

Struktur html detik.com memiliki 2 jenis pola peletakan judul berita, yaitu judul berita video dan judul berita narasi. Berita video dan berita narasi memiliki struktur penulisan judul yang berbeda pada halamannya, berikut analisis kedua jenis berita tersebut.

➤ Berita Video

```

▼ <div class="headline hldetail">
  ▼ <article>
    ▶ <span class="ratiobox ratio_16_9">...</span>
    ▼ <div class="text_area">
      <span class="label">detikFlash</span>
      <h1>Hore! Tol Fungsional Salatiga-Kartasura Bisa
      Dipakai saat Mudik</h1> == $0
      ▶ <span class="desc">...</span>
      <span class="date">Eko Susanto - 20DETIK </span>
      <span class="date">Rabu, 06 Jun 2018 08:00 WIB</span>
    </div>
  </article>

```

Gambar 4. 4 Struktur halaman berita video

Gambar 4.4 menggambarkan struktur penulisan judul berita pada halaman berita video situs www.detik.com. Penulisan judul pada halaman berita video di situs www.detik.com secara umum adalah <article class="teks_area"><h1>Judul Berita Video</h1></article>. Pola penulisan ini hanya digunakan untuk menuliskan judul berita, sehingga dapat dijadikan rujukan untuk mendapatkan konten judul berita.

➤ Berita Narasi / Tekstual

```

<!--S:JUDUL ARTICLE-->
▼ <div class="detail_area">
  <div class="date">Rabu 06 Juni 2018, 21:25 WIB</div>
  <div class="sub"></div>
  <h1 class="jdl">Percobaan Alarm Kebakaran Bikin Panik
  Pengunjung Mal di Bandung</h1> == $0
  ▶ <div class="author">...</div>
  <div class="clearfix"></div>
</div>

```

Gambar 4. 5 Struktur html halaman berita narasi/tekstual

Gambar 4.5 menggambarkan struktur html halaman berita narasi/tekstual pada situs www.detik.com. Secara umum pola penulisan judul berita pada situs

www.detik.com adalah `<h1 class="jdl">Judul Berita</h1>`. Pola penulisan ini hanya digunakan untuk menuliskan judul berita, sehingga dapat dijadikan rujukan untuk mendapatkan konten judul berita.

2. Liputan 6

Struktur html liputan6.com menunjukkan terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur penulisan judul yang sama. Berikut adalah analisis struktur html halaman berita www.liputan6.com.

```
▼ <div class="article-header" data-component="mobile:article-header" data-component-name="mobile:article-header">
  <h1 class="article-header__title">VIDEO: Jelang Lebaran, Pemerintah Pulangkan 404 WNI dari Malaysia</h1> == $0
  ▼ <span class="article-header__authors">
    " Oleh "
    ▼ <a class="article-header__author-link" href="https://m.liputan6.com/me/108792894144497515645">
      "Gautama Adianto"
      ::after
    </a>
  </span>
  <span class="article-header__datetime"> pada 07 Jun 2018, 08:30 WIB</span>
</div>
```

Gambar 4. 6 Struktur halaman berita liputan6.com

Gambar 4.6 menggambarkan struktur html halaman berita narasi/tekstual pada situs www.liputan6.com. Secara umum pola penulisan judul berita pada situs www.liputan6.com adalah `<h1 class="article-header__title">Judul Berita</h1>`. Pola penulisan ini hanya digunakan untuk menuliskan judul berita, sehingga dapat dijadikan rujukan untuk mendapatkan konten judul berita.

3. Tribunnews

Struktur html tribunnews.com menunjukkan terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur penulisan judul yang sama. Berikut adalah analisis struktur html halaman berita www.tribunnews.com.

Gambar 4.8 menggambarkan struktur html halaman berita video dari situs www.detik.com. Secara umum pola penulisan isi berita pada berita video di situs www.detik.com adalah `Deskripsi Video / Isi Berita`. Pola ini hanya digunakan untuk menuliskan deskripsi dari video yang ditayangkan, sehingga dapat dijadikan acuan pembuatan pola untuk mendapatkan konten berupa isi berita dari halaman berita.

➤ Berita Narasi / Tekstual

```

▼ <div class="text_detail detail_area" id="detikdetailtext" == $0
  <strong>Jakarta </strong>
  "-
  Partai Gerindra menyayangkan kebijakan tunjangan hari raya ("
  <a href="https://www.detik.com/tag/thr/" target="_blank">THR</a>
  ") untuk PNS daerah yang dibebankan ke APBD. Meski demikian, Gerindra tak heran dengan
  kebijakan pemerintah pusat yang diterapkan pada tahun politik ini."
  <br>
  <br>
  ""Tahulah kita, ini kan memang tahun politik," ujar Wakil Ketua F-Gerindra Ahmad Riza
  Patria kepada wartawan, Rabu (6/6/2018). "
  <br>
  <br>
  ▶ <table class="linksisip">...</table>
  <br>
  "Menurut Riza, kebijakan THR tahun ini tak selaras dengan pernyataan pemerintah terkait
  kondisi keuangan negara. Riza lalu mempertanyakan asal anggaran THR. "
  <br>
  <br>
  <!--s:parallaxindetail-->
  ▶ <center>...</center>
  <!--e:parallaxindetail-->
  ""Pemerintah selama ini kan bahasanya tidak punya uang, defisit anggaran, untuk honorer
  mengangkat tidak punya uang, rata-rata tidak ada uang. Tiba-tiba ada uang jumlahnya
  hampir Rp 36 T. Pertanyaannya, dari mana uangnya?" kata Riza. "
  <br>
  <br>
  ▶ <table align=""center"" class=""pic_artikel_sisip_table"">...</table>
  <br>
  "Bagi Riza, kebijakan membebaskan anggaran "
  <a href="https://www.detik.com/tag/thr-pns/" target="_blank">THR PNS</a>
  " daerah ke APBD tak tepat. Malah, menurut dia, pemerintah memaksakan pemberian THR."
  <br>
  <br>
  ""Memang memaksa karena tidak dibahas sebelumnya, tidak didiskusikan sebelumnya, tidak
  dianggarkan sebelumnya, apalagi angkanya sangat signifikan peningkatannya. Kan daerah-
  daerah kan banyak yang tidak siap. Ini kan menimbulkan masalah baru," kritik Ketua DPP "
  <a href="https://www.detik.com/tag/gerindra/" target="_blank">Gerindra</a>
  " itu.
  "
  <strong>(gbr/jbr)</strong>

```

Gambar 4. 9 Struktur html halaman berita narasi/tekstual

Gambar 4.9 menggambarkan struktur html halaman berita narasi/tekstual pada situs www.detik.com. Isi berita pada struktur halaman ini secara umum adalah `<div class="teks_detail detail_area" id="detikdetailteks">Isi Berita Narasi</div>`. Pola ini hanya digunakan untuk menuliskan isi berita sehingga dapat dijadikan acuan untuk pembuatan pola untuk mendapatkan konten berupa isi berita.

Sebelum melakukan ekstraksi isi berita, sisipan-sisipan konten yang bukan merupakan bagian dari isi berita harus dihilangkan terlebih dahulu. Ada 3 jenis

sisipan konten yang disisipkan pada konten berita, yaitu sisipan iklan, tautan untuk berita lain, dan sisipan gambar. Berikut adalah hasil analisis untuk menghilangkan konten sisipan pada isi berita.

1. Sisipan iklan

Struktur pada Gambar 4.8 menunjukkan adanya sisipan iklan yang diletakkan pada isi berita. Sisipan iklan ini diletakkan di antara tag `<!--s:parallaxindetail-->` dan `<!--e:parallaxindetail`. Selain menggunakan pola penulisan itu, sisipan iklan pada detik.com, terkadang ditulis menggunakan tag `<blockquote></blockquote>`.

2. Sisipan tautan lain

Struktur pada Gambar 4.8 menunjukkan adanya sisipan tautan yang diletakkan pada isi berita. Sisipan tautan ini diletakkan di antara tag `<table class="tautansisip">` dan `</table>`. Selain menggunakan pola itu, terkadang tautan lain ditulis menggunakan tag `<a>`.

3. Sisipan gambar

Struktur pada Gambar 4.8 menunjukkan adanya sisipan gambar yang diletakkan pada isi berita. Sisipan gambar ini diletakkan di antara tag `<table align="center" class="pic_article_sisip_table">` dan `</table>`. Selain menggunakan pola penulisan tersebut, gambar pada halaman ini juga terkadang dituliskan menggunakan dua pola penulisan lain, yaitu di antara tag `<iframe></iframe>` dan menggunakan tag ``.

Setelah semua sisipan non-konten terhapus dari isi berita, maka proses ekstraksi berita dapat dilakukan. Terkadang, selain berisi tag html untuk pengaturan paragraf, disisipkan juga tag script di dalam berita. Maka dari itu, sebelum disimpan semua tag html ataupun tag script yang berada pada hasil ekstraksi harus dihapus terlebih dahulu.

2. Liputan 6

Struktur html liputan6.com menunjukkan terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur isi berita yang sama. Berikut analisis struktur html halaman berita www.liputan6.com.

```

▼<div class="article-content-body__item">
  ▼<div class="article-raw-content" itemprop="description" data-component-name="mobile:article-raw-content">
    ▼<p>
      <b>Liputan6.com, Jakarta</b>
      " Kementerian Luar Negeri pulangkan 404 WNI kelompok rentan dari detensi imigrasi Malaysia dan dari penampungan sementara KBRI Kuala Lumpur serta KJRI Johor Bahru ." == $0
    </p>
  </div>
</span class="article-content-body__item-overlay"></span>
</div>
</div>

```

Gambar 4. 10 Struktur halaman berita situs liputan6.com

Gambar 4.10 menggambarkan struktur html halaman berita dari situs www.liputan6.com. Struktur html pada gambar 4.10 menunjukkan bahwa isi berita diletakkan di antara tag `<div></div>`. Isi berita pada situs ini secara umum adalah `<div class="article-raw-content" itemprop="description" data-component-name="mobile:article-raw-content">Isi Berita</div>`.

Sebelum melakukan ekstraksi isi berita, sisipan-sisipan konten yang bukan merupakan bagian dari isi berita harus dihilangkan terlebih dahulu. Ada 3 jenis sisipan konten yang disisipkan pada konten berita, yaitu sisipan iklan, tautan untuk berita lain, dan sisipan gambar. Berikut adalah hasil analisis untuk menghilangkan konten sisipan pada isi berita.

1. Sisipan iklan

Sisipan iklan pada halaman ini diletakkan di antara tag `<div class="seamless-ads__container">` dan `</div>`. Selain menggunakan pola penulisan itu, sisipan iklan pada liputan6.com, terkadang ditulis menggunakan tag `<blockquote>`.

2. Sisipan tautan lain

Sisipan tautan pada halaman ini diletakkan di antara tag `<div class="seamless-ads">` dan `</div>`. Selain menggunakan pola itu, terkadang tautan lain ditulis menggunakan tag `<a>` atau di antara tag `<div class="baca-juga"></div>`.

3. Sisipan gambar

Sisipan gambar pada halaman ini diletakkan di antara menggunakan tag ``. Setelah sisipan gambar berhasil di ekstraksi, hasil ekstraksi tersebut di hapus

dari isi berita.

Setelah semua sisipan non-konten terhapus dari isi berita, maka proses ekstraksi berita dapat dilakukan. Terkadang, selain berisi *tag* html untuk pengaturan paragraf, disisipkan juga *tag* `<iframe>` yang memuat video berita lain di dalam berita. Maka dari itu, sebelum disimpan semua *tag* html ataupun sisipan video lain yang berada pada hasil ekstraksi harus dihapus terlebih dahulu.

3. Tribunnews

Struktur html Tribunnews.com menunjukkan terdapat dua jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur isi berita yang sama. Berikut analisis struktur html halaman berita www.liputan6.com.

```

▼<div class="txt-article mb20">
  ▶<p>...</p>
  ▶<p>...</p>
  ▶<p>...</p>
  ▶<p>...</p>
  <span id="innity-in-post" style="padding:0px"></span>
  ▶<div id="div-Inside-MediumRectangle" style="text-align: center; margin: auto;" data-google-query-id="
  "CLmejNv989sCFZg_KwodFtkCxw">...</div>
  <p>"Saya akui tim Pak Dirman hebat. Kita kalah saing di sana," ucapnya.</p>
  ▶<p class="baca" style>...</p>
  ▶<p>...</p>
  <script src="//a.teads.tv/page/73726/tag" async="true"></script>
  <p>Di ibu kota Jawa Tengah, Ganjar meraup kemenangan hingga 71 persen.</p>
  <div class style="padding:0px!important"></div>
  ▶<p>...</p>
  ▶<p>...</p>
  <p>Menurut dia, berdasarkan hitungan kemenangan terbawah diangka 61,6 persen.</p>
  ▶<p>...</p>
  <p>Pihaknya menduga kekurangan 2 persen terjadi di Tegal, Brebes dan Kebumen.</p>
  ▶<p>...</p>
  ▶<p>...</p>
  ▶<p>...</p>
  <p>Ada 400 TPS yang dijadikan sampel dengan margin eror +/- 1 persen.</p>
  <p>Partisipasi pemilih mencapai 67,45 persen.</p>
  ▶<p>...</p>
</div>
<div id="fixsharebottom"></div>
▶<div class="pa15">...</div>
<div class="cl2"></div>
</div>

```

Gambar 4. 11 Struktur halaman berita www.tribunnews.com

Gambar 4.11 menggambarkan struktur html halaman berita pada situs www.tribunnews.com. Struktur halaman ini menunjukkan bahwa isi berita diletakkan pada *tag* `<div>` Isi berita pada struktur halaman ini secara umum adalah `<div class="teks-article mb20">Isi Berita </div>`. Pola ini hanya digunakan untuk menuliskan isi berita sehingga dapat dijadikan acuan untuk pembuatan pola untuk mendapatkan konten berupa isi berita.

Sebelum melakukan ekstraksi isi berita, sisipan-sisipan konten yang bukan merupakan bagian dari isi berita harus dihilangkan terlebih dahulu. Ada empat jenis

sisipan konten yang disisipkan pada konten berita, yaitu sisipan *tag* ``, tautan untuk berita lain, sisipan iklan dan sisipan *script*. Berikut adalah analisis untuk menghilangkan konten sisipan pada isi berita.

1. Sisipan *tag* ``

Struktur pada Gambar 4.11 menunjukkan adanya sisipan *tag* `` yang diletakkan pada isi berita. Secara umum penulisan sisipan *tag* ini adalah ``.

2. Sisipan iklan

Struktur pada Gambar 4.11 menunjukkan adanya sisipan iklan yang diletakkan pada isi berita. Sisipan iklan ini diletakkan di antara *tag* `<div></div>`. Pola penulisan sisipan iklan secara umum pada halaman berita www.tribunnews.com adalah `<div id="div-Inside-MediumRectangle" style="teks-align: center; margin: auto;" data-google-query-id="COyBwKPS9NsCFQcKjgodX7Q01A">Isi Iklan</div>`. Selain menggunakan pola penulisan itu, sisipan iklan pada halaman ini terkadang ditulis di antara *tag* `<div class='adspruce-bannerspot'></div>`.

3. Sisipan tautan lain

Struktur pada Gambar 4.11 menunjukkan adanya sisipan tautan yang diletakkan pada isi berita. Sisipan tautan ini diletakkan di antara *tag* `<p></p>`. Pola penulisan *tautan* berita lain pada halaman www.tribunnews.com adalah `<p class="baca"> Baca :judul berita</p>`. Selain menggunakan pola penulisan itu, tautan berita lain terkadang ditulis menggunakan *tag* `` dan *tag* `<a>`.

4. Sisipan *script*

Struktur pada Gambar 4.11 menunjukkan adanya sisipan *script* yang diletakkan pada isi berita. Sisipan *script* ini diletakkan di antara *tag* `<script></script>`. Pola penulisan sisipan gambar pada halaman berita www.tribunnews.com adalah `<script src="Sumber Script" async="true"></script>`.

5. Sisipan Gambar

Sisipan gambar pada halaman ini diletakkan di antara menggunakan *tag* `<figure>`. Setelah sisipan gambar berhasil di ekstraksi, hasil ekstraksi tersebut di hapus dari isi berita.

Setelah semua sisipan non-konten terhapus dari isi berita, maka proses ekstraksi berita dapat dilakukan. Hasil ekstraksi isi berita akan berupa teks dalam bentuk paragraf beserta *tag* html yang berfungsi untuk pengaturannya. Sebelum disimpan semua *tag* html yang berada pada hasil ekstraksi harus dihapus terlebih dahulu.

4.1.4. Observasi Penulisan Penulis Berita

1. Detik

Analisis struktur html detik.com menunjukkan bahwa terdapat 2 jenis pola peletakan penulis / sumber berita, yaitu penulis/sumber berita video dan penulis/sumber berita narasi. Berita video dan berita narasi memiliki struktur penulisan penulis/sumber yang berbeda pada halamannya, berikut hasil analisis kedua jenis berita tersebut.

➤ Berita Video

```

▼ <article>
  ▶ <span class="ratiobox ratio_16_9">...</span>
  ▼ <div class="text_area">
    <span class="label">Spesial Ramadan</span>
    <h1>Bahtera Nabi Nuh dan Berkah Minyak di Laut Kaspia</h1>
    ▼ <span class="desc">
      ▶ <p>...</p>
      </span>
      <span class="date">Muhammad Abdurrosyid - 20DETIK</span> == $0
      <span class="date">Rabu, 06 Jun 2018 19:14 WIB</span>
    </div>
  </article>

```

Gambar 4. 12 Struktur html halaman berita video detik.com

Gambar 4.12 menggambarkan struktur html halaman berita video dari situs www.detik.com. Secara umum pola penulisan penulis/sumber berita pada situs www.detik.com yang dapat dijadikan acuan adalah Nama Penulis/Sumber Berita.

➤ Berita Narasi / Tekstual


```

<!--S:JUDUL ARTICLE-->
▼<div class="detail_area">
  <div class="date">Rabu 06 Juni 2018, 21:25 WIB</div>
  <div class="sub"></div>
  <h1 class="jdl">Percobaan Alarm Kebakaran Bikin Panik
  Pengunjung Mal di Bandung</h1>
  ▼<div class="author">
    <strong>Dony Indra Ramadhan</strong> == $0
    " - detikNews
    "
  </div>
  <div class="clearfix"></div>
</div>

```

Gambar 4. 13 Struktur html halaman berita narasi/tekstual detik.com

Gambar 4.13 menggambarkan struktur html halaman berita narasi/tekstual dari situs www.detik.com. Struktur html pada gambar 4.13 menunjukkan bahwa penulis/sumber berita yang berjenis berita narasi/tekstual diletakkan di antara *tag* `<div></div>`. Secara umum pola penulisan penulis/sumber berita pada situs www.detik.com adalah `<div class="author">Nama Penulis</div>`.

Hasil ekstraksi penulis/sumber berita akan mengandung *tag* ``. Sebelum data disimpan ke dalam basis data, maka *tag* `` harus dihapus terlebih dahulu.

2. Liputan 6

Struktur html liputan6.com menunjukkan bahwa terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur penulisan penulis/sumber yang sama. Berikut adalah analisis struktur html yang digunakan.

```

▼<div class="article-header" data-component="mobile:article-header" data-component-
name="mobile:article-header">
  <h1 class="article-header__title">VIDEO: Jelang Lebaran, Pemerintah Pulangkan
  404 WNI dari Malaysia</h1>
  ▼<span class="article-header__authors">
    " Oleh "
    ▼<a class="article-header__author-link" href="https://m.liputan6.com/me/
    108792894144497515645">
      "Gautama Adianto" == $0
      ::after
    </a>
  </span>
  <span class="article-header__datetime"> pada 07 Jun 2018, 08:30 WIB</span>
</div>

```

Gambar 4. 14 Struktur html halaman berita liputan6.com

Gambar 4.14 menggambarkan struktur html halaman berita dari situs www.liputan6.com. Struktur html pada gambar 4.14 menunjukkan bahwa penulis/sumber berita diletakkan di antara *tag* `<a>`. Pola penulisan nama

penulis/sumber berita pada situs www.liputan6.com secara umum adalah `Nama Penulis/Sumber`. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil nama penulis/sumber berita.

3. Tribunnews

Struktur html tribunnews.com menunjukkan bahwa terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur penulisan penulis/sumber yang sama. Berikut adalah analisis struktur html yang digunakan.

```
▼ <div class="f12 grey mb15">
  <div>Penulis: Srihandriatmo Malau </div>
</div>
```

Gambar 4. 15 Struktur penulisan nama penulis pada www.tribunnews.com

Gambar 4.15 menggambarkan struktur html halaman berita dari situs www.tribunnews.com. Struktur html pada gambar 4.15 menunjukkan bahwa penulis/sumber berita diletakkan di antara tag `<div></div>`. Pola penulisan nama penulis/sumber berita pada situs www.tribunnews.com secara umum adalah `<div class="f12 grey mb15"> Nama Penulis/Sumber</div>`. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil nama penulis/sumber berita. Hasil dari proses ekstraksi penulis akan mengandung tag html yang berfungsi untuk mengatur tampilan nama penulis tersebut, sehingga sebelum nama penulis disimpan ke dalam *basis data* maka tag-tag html yang ikut terekstrak harus di hapus terlebih dahulu.

4.1.5. Observasi Penulisan Tanggal Publikasi Berita

1. Detik

Analisis struktur html detik.com menunjukkan terdapat 2 jenis pola peletakan tanggal publikasi berita yang ditemukan, yaitu tanggal publikasi berita video dan tanggal publikasi berita narasi. Berita video dan berita narasi memiliki struktur penulisan tanggal publikasi yang berbeda pada halamannya, berikut hasil analisis kedua jenis berita tersebut.

➤ Berita Video

```

▼<article>
  ▶<span class="ratiobox ratio_16_9">...</span>
  ▼<div class="text_area">
    <span class="label">Special Ramadan</span>
    <h1>Bahtera Nabi Nuh dan Berkah Minyak di Laut Kaspia
    </h1>
    ▼<span class="desc">
      ▶<p>...</p>
      </span>
      <span class="date">Muhammad Abdurrosyid - 20DETIK
      </span> == $0
      <span class="date">Rabu, 06 Jun 2018 19:14 WIB</span>
    </div>
  </article>

```

Gambar 4. 16 Struktur html halaman berita video detik.com

Gambar 4.16 menggambarkan struktur html halaman berita video dari situs www.detik.com. Struktur html pada gambar 4.16 menunjukkan bahwa tanggal publikasi berita yang berjenis berita video diletakkan di antara *tag* ``. Secara umum pola penulisan penulis/sumber berita pada situs www.detik.com adalah `20DETIK Tanggal Publikasi`.

➤ Berita Narasi / Tekstual

```

<!--S:JUDUL ARTICLE-->
▼<div class="detail_area">
  <div class="date">Rabu 06 Juni 2018, 19:16 WIB</div> ==
  <div class="sub"></div>
  <h1 class="jd1">Gerindra soal THR PNS Daerah: Tahulah
  Kita Ini Tahun Politik</h1>
  ▶<div class="author">...</div>
  <div class="clearfix"></div>
</div>

```

Gambar 4. 17 struktur html halaman berita narasi/tekstual detik.com

Gambar 4.17 menggambarkan struktur html halaman berita narasi/tekstual pada situs www.detik.com. Struktur html pada gambar 4.17 menunjukkan bahwa tanggal publikasi berita diletakkan *di antara tag* `<div></div>`. Pola penulisan tanggal publikasi berita pada situs www.detik.com secara umum adalah `<div class="date">Tanggal Publikasi Berita</div>`. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil tanggal publikasi berita. Hasil dari proses ekstraksi tanggal publikasi akan mengandung *tag* html yang berfungsi untuk mengatur tampilan tanggal publikasi tersebut, sehingga sebelum tanggal publikasi disimpan ke dalam *basis data* maka *tag-tag* html yang ikut terekstrak harus di hapus terlebih dahulu.

2. Liputan 6

Analisis struktur html liputan6.com menunjukkan terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur tanggal publikasi yang sama. Berikut adalah analisis struktur halaman berita yang digunakan.

```
<div class="article-header" data-component="mobile:article-header" data-component-name="mobile:article-header">
  <h1 class="article-header__title">VIDEO: Jelang Lebaran, Pemerintah Pulangkan
  404 WNI dari Malaysia</h1>
  <span class="article-header__authors">
    " Oleh "
    <a class="article-header__author-link" href="https://m.liputan6.com/me/
    108792894144497515645">
      "Gautama Adianto"
      ::after
    </a>
  </span>
  <span class="article-header__datetime"> pada 07 Jun 2018, 08:30 WIB</span> == $0
</div>
```

Gambar 4. 18 Struktur html halaman berita liputan6.com

Gambar 4.18 menggambarkan struktur html halaman berita dari situs www.liputan6.com. Struktur html pada gambar 4.18 menunjukkan bahwa tanggal publikasi berita diletakkan di antara *tag* ``. Pola penulisan tanggal publikasi berita pada situs www.liputan6.com secara umum adalah `Tanggal Publikasi Berita`. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil tanggal publikasi berita. Hasil dari proses ekstraksi tanggal publikasi akan mengandung *tag* html yang berfungsi untuk mengatur tampilan tanggal publikasi tersebut, sehingga sebelum tanggal publikasi disimpan ke dalam *basis data* maka *tag-tag* html yang ikut terekstrak harus di hapus terlebih dahulu.

3. Tribunnews

Analisis struktur html tribunnews.com menunjukkan terdapat 2 jenis berita yaitu berita video dan narasi/tekstual. Dua jenis berita tersebut memiliki struktur tanggal publikasi yang sama. Berikut adalah analisis struktur halaman berita yang digunakan.

```

<h1 class="f32 fno crimson" style="line-height:40px;font-size:
34px">Hitung Cepat LSI Pilgub Sumsel: Herman Deru-Mawardi
Yahya Unggul 81,04 persen</h1>
▶ <h4 class="hide">...</h4>
▼ <div class="mt10 mb10">
  <time class="grey f13 dip">Rabu, 27 Juni 2018 15:16 WIB
  </time>
  <div class="grey bdr3 c12 mt10"></div>
</div>

```

Gambar 4. 19 Struktur Penulisan Tanggal Publikasi www.tribunnews.com

Gambar 4.19 menggambarkan struktur html halaman berita dari situs www.liputan6.com. Struktur html pada gambar 4.19 menunjukkan bahwa tanggal publikasi berita diletakkan *di antara tag* `<time></time>`. Pola penulisan tanggal publikasi berita pada situs www.tribunnews.com secara umum adalah `<time class="grey f13 dip">Tanggal Publikasi Berita</time>`. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil tanggal publikasi berita. Hasil dari proses esktraksi tanggal publikasi akan mengandung *tag* html yang berfungsi untuk mengatur tampilan tanggal publikasi tersebut, sehingga sebelum tanggal publikasi disimpan ke dalam *basis data* maka *tag-tag* html yang ikut terekstrak harus di hapus terlebih dahulu.

4.1.6. Observasi Penulisan Tautan Berita Berupa Video

1. Detik

Tautan berita berupa video hanya akan dicari dan di ekstrak ketika berita yang ditampilkan merupakan jenis berita video. Pada situs www.detik.com jenis berita video atau bukan dapat dilihat pada *tautan* dari berita itu sendiri. Secara umum pola penulisan *tautan* berita video pada detik adalah <https://m.detik.com/20detik/detikflash/video-id/judul-berita> atau <https://m.detik.com/news/detiktv/video-id/judul-berita>. Berdasarkan dua model *tautan* tersebut dapat disimpulkan bahwa pengambilan tautan video hanya akan dilakukan jika *tautan* berita mengandung salah satu dari kata 20detik, detikflash atau detiktv. Berikut adalah analisis halaman berita berjenis berita video pada situs www.detik.com.


```

▼ <article> == $0
  ▼ <span class="ratiobox ratio_16_9">
    ▼ <span class="ratiobox_content">
      ▶ <iframe src="https://m.detik.com/20detik/embed/180627077?
        autostart=1" frameborder="0" width="420" scrolling="no"
        allowfullscreen="true" height="236">...</iframe>
      </span>
      ::after
    </span>
    ▶ <div class="text_area">...</div>
  </article>

```

Gambar 4. 20 Struktur penulisan tautan berita video detik.com

Gambar 4.20 menggambarkan struktur html halaman berita video dari situs www.detik.com. Struktur html pada gambar 4.20 menunjukkan bahwa tautan video berita diletakkan *di antara tag* `<iframe></iframe>`. Pola penulisan tautan video pada situs www.detik.com secara umum adalah `<iframe src="tautan video" frameborder="0" width="420" scrolling="no" allowfullscreen="true" height="236"></iframe>`. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil tautan video pada berita.

2. Liputan 6

Halaman berita pada situs liputan6.com selalu menyisipkan tautan video di dalamnya. Sisipan video tidak hanya dilakukan pada jenis berita video saja, namun juga pada struktur penulisan berita narasi. Pada berita narasi, tautan video yang disisipkan adalah tautan video berita lain, sedangkan pada berita video, tautan video yang disisipkan adalah tautan video untuk berita video tersebut. Pola yang digunakan untuk menampilkan sisipan video pada berita narasi dan berita video dapat dilihat pada gambar 4.21 dan 4.22.

```

▶ <iframe class="vidio-embed article-video-gallery--
  item_vidio-embed" src="https://www.vidio.com/embed/1406729-
  asal-usul-nama-daerah-di-jakarta-t...i?source=liputan6-
  news&medium=videogallery&autoplay=false&player_only=true"
  width="100%" height="205" frameborder="0" scrolling
  allowfullscreen data-component-name="ui:iframe:vidio">...
</iframe>

```

Gambar 4. 21 Struktur penulisan tautan video berita video www.liputan6.com

```

▶<iframe class="vidio-embed article-video-gallery--
item_vidio-embed" src="https://www.vidio.com/embed/
1372290-ketum-ppp-saya-siap-dampingi-jokow...us-pagi?
source=liputan6-
news&medium=embed&autoplay=false&player_only=false"
width="100%" height="205" frameborder="0" scrolling
allowfullscreen data-component-name="ui:iframe:vidio">
</iframe>

```

Gambar 4. 22 Struktur penulisan tautan video berita narasi www.liputan6.com

Gambar 4.21 dan 4.22 menggambarkan struktur penulisan tautan video pada dua jenis berita. Semua elemen html berupa atribut dan *value* yang digunakan pada kedua jenis berita tersebut sama persis. Perbedaan tautan pada kedua berita hanya terletak pada tautan video yang ditautkan pada kedua jenis berita tersebut. Pada jenis berita video, atribut *player_only* pada tautan bernilai *true*, sedangkan pada berita narasi nilai atribut tersebut adalah *false*. Hal ini dapat dijadikan acuan untuk membedakan tautan video untuk berita video dan tautan video untuk berita narasi.

3. Tribunnews

Tribunnews mempunyai dua jenis berita, yaitu berita video dan narasi. Perbedaan kedua jenis berita ini hanya terletak pada ada atau tidaknya tautan video yang disisipkan ke dalamnya. Berikut adalah analisis halaman berita berjenis berita video pada situs www.tribunnews.com.

```

▼<div class="video" id="iframe-resize">
▶<iframe width="300" height="169" src="http://www.youtube-
nocookie.com/embed/HH0ckw07PgA" frameborder="0" allowfullscreen
style="width: 100%; height: 202.8px;"></iframe>
</div>

```

Gambar 4. 23 Struktur Penulisan Tautan Video pada www.tribunnews.com

Gambar 4.23 menggambarkan struktur html halaman berita video dari situs www.tribunnews.com. Struktur html pada gambar 4.23 menunjukkan bahwa tautan video berita diletakkan *di antara tag* <iframe></iframe>. Pola penulisan tautan video pada situs www.tribunnews.com secara umum adalah <div class="video" id="frame-resize"><iframe width="300" height="169" src="http://www.youtube-nocookie.com/embed/HH0ckw07PgA" frameborder="0" allowfullscreen="" style="width: 100%; height: 202.8px;"></iframe>. Pola ini dapat dijadikan acuan untuk mendapatkan pola *Regex* untuk mengambil tautan video pada berita.

4.2. Pembentukan Pola Ekspresi Reguler

Pola *regex* dibentuk berdasarkan pola penulisan masing-masing konten yang ditemukan pada tahap analisis struktur berita. Penyusunan pola *regex* menggunakan acuan *tag* apa yang digunakan untuk menampilkan konten, atribut dari *tag* yang digunakan, dan elemen lain yang mendukung agar pola *regex* yang dibentuk bersifat unik dan hanya digunakan untuk menampilkan konten yang bersangkutan. Berikut adalah proses pembentukan pola *regex*.

4.2.1. Pola *Regex Tautan* Berita

1. Detik

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tautan / *tautan* berita pada halaman indeks yang dapat dijadikan acuan adalah `<a data-category="Subkanal detikNews" data-cation="Indeks" data-label="List Berita" href="Tautan Menuju Halaman Berita" class="list">Judul Berita Yang Dituju`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dari *tag* `<a>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak *tautan* dari daftar berita yang ditampilkan.

2. Liputan 6

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tautan / *tautan* berita pada halaman indeks yang dapat dijadikan acuan adalah `</div>Judul Berita Yang Dituju`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* penutup `</div>` dan *tag* pembuka dari *tag* `<a>` beserta atribut *title* yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak *tautan* dari daftar berita yang ditampilkan.

3. Tribunnews

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tautan / *tautan* berita pada halaman indeks yang dapat dijadikan acuan adalah `<h3 class="blue"><a href="Tautan Menuju Halaman`

Berita" title="Judul Berita" class="fbo2 f16 ln22 txt-oev-3">Judul Berita Yang Dituju</h3>. Berdasarkan pola penulisan tersebut, maka dibutuhkan dua pola *regex* untuk melakukan ekstraksi.

Pola *regex* yang pertama berfungsi untuk mencari isi dari tag <h3> yang memiliki *class* dengan *value* "blue". Pola ini dibentuk berdasarkan hasil analisis yang menunjukkan bahwa tag <h3 class="blue"> hanya digunakan untuk menampilkan judul dan tautan berita saja, sehingga dapat dijadikan acuan bahwa, tag <a> yang ditemukan di dalam tag <h3 class="blue"> adalah tag <a> yang mengandung tautan menuju halaman berita.

Pola *regex* yang kedua berfungsi untuk mencari tag <a> yang ada pada hasil ekstraksi pola pertama (yang berada di dalam tag <h3 class="blue"></h3>). Pola ini dibentuk untuk mengekstrak tautan berita yang ada pada atribut *href* milik tag <a> yang ditemukan.

4.2.2. Pola *Regex* Judul Berita

1. Detik

Struktur html detik.com memiliki dua jenis pola peletakan judul berita, yaitu judul berita video dan judul berita narasi. Berdasarkan hasil analisis struktur halaman dari dua jenis berita tersebut, berikut adalah proses pembentukan pola *regex* yang dibutuhkan.

➤ Berita Video.

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan judul berita pada halaman berita video yang dapat dijadikan acuan adalah <article class="teks_area"><h1>Judul Berita Video</h1></article>. Berdasarkan pola penulisan tersebut, maka dengan menggunakan tag pembuka dan penutup dari tag <h1> beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

➤ Berita Narasi / Tekstual

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan judul berita pada halaman berita narasi yang dapat dijadikan acuan adalah <h1 class="jd1">Judul Berita</h1>. Berdasarkan pola

penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<h1>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

2. Liputan 6

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan judul berita pada halaman berita video dan narasi yang dapat dijadikan acuan adalah `<h1 class="article-header__title">Judul Berita</h1>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<h1>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

3. Tribunnews

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan judul berita pada halaman berita video dan narasi yang dapat dijadikan acuan adalah `<h1 class="f32 fno crimson" style="line-height:40px;font-size:34px">Judul Berita</h1>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<h1>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

4.2.3. Pola *Regex* Isi Berita

1. Detik

Struktur html detik.com memiliki dua jenis pola peletakan isi berita, yaitu isi berita video dan isi berita narasi. Berdasarkan hasil analisis struktur halaman dari dua jenis berita tersebut, berikut adalah proses pembentukan pola *regex* yang dibutuhkan.

➤ Berita Video

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan isi berita pada halaman berita video yang dapat dijadikan acuan adalah `Deskripsi Video / Isi Berita`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak isi berita yang ditampilkan.

Deskripsi berita video yang berhasil di ekstrak terkadang memiliki tautan ke berita lain, yang ditulis di dalam tag `<a>`, maka dari itu perlu ditambahkan pola *regex* yang berfungsi untuk menghapus tautan tersebut dari dalam berita. Selain itu, dibutuhkan juga pola *regex* tambahan untuk menghapus *tag-tag* html yang masih terdapat pada hasil ekstraksi berita.

➤ Berita Narasi / Tekstual

Hasil analisis struktur halaman ini menunjukkan bahwa isi berita diletakkan pada tag `<div>` dengan atribut *class* yang memiliki *value* “teks_detail detail_area” dan *id* yang memiliki *value* “detikdetailteks”, sehingga isi berita pada struktur halaman ini secara umum adalah `<div class="teks_detail detail_area" id="detikdetailteks">Isi Berita Narasi</div>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan tag pembuka dan penutup dari tag `<div>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak isi berita yang ditampilkan.

Untuk membersihkan berita dari konten non-berita, dibutuhkan pola *regex* tersendiri untuk menemukan kemudian menghapusnya. Berikut adalah proses pembentukan pola *regex* tersebut.

1. Sisipan iklan

Sisipan iklan pertama diletakkan di antara tag `<!--s:parallaxindetail-->` dan `<!--e:parallaxindetail-->`. Pola *regex* yang dibentuk cukup disusun menggunakan tag pembuka dan penutup dari tag ini. Sisipan iklan kedua ditulis menggunakan tag `<blockquote></blockquote>`. Pola *regex* yang dibentuk cukup disusun menggunakan tag pembuka dan penutup dari tag ini.

2. Sisipan tautan lain

Sisipan tautan pertama diletakkan di antara tag `<table class="tautansisip">` dan `</table>`. Pola *regex* yang dibentuk cukup disusun menggunakan tag pembuka dan penutup dari tag ini. Sisipan tautan kedua ditulis menggunakan tag `<a>`. Pola *regex* yang dibentuk cukup disusun menggunakan tag pembuka dari tag ini.

3. Sisipan gambar

Sisipan gambar pertama diletakkan di antara tag `<table align="center"`

`class="pic_article_sisip_table">` dan `</table>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup dari *tag* ini. Sisipan gambar kedua ditulis menggunakan *tag* ``. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dari *tag* ini.

2. Liputan 6

Hasil analisis struktur halaman ini menunjukkan bahwa pola penulisan isi berita yang dapat dijadikan acuan pembuatan pola *regex* adalah `<div class="article-raw-content" itemprop="description" data-component-name="mobile:article-raw-content">Isi Berita</div>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<div>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak isi berita yang ditampilkan.

Untuk membersihkan berita dari konten non-berita, dibutuhkan pola *regex* tersendiri untuk menemukan kemudian menghapusnya. Berikut adalah proses pembentukan pola *regex* tersebut.

1. Sisipan iklan

Sisipan iklan pertama diletakkan *di antara tag* `<div class="seamless-ads__container"></div>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup dari *tag* ini. Sisipan iklan kedua ditulis menggunakan *tag* `<blockquote>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka beserta atribut yang dipakai *tag* ini.

2. Sisipan tautan lain

Sisipan tautan pertama diletakkan *di antara tag* `<div class="seamless-ads">` dan `</div>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup dari *tag* ini. Sisipan tautan kedua ditulis menggunakan *tag* `<a>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dari *tag* ini. Sisipan tautan ketiga ditulis menggunakan *tag* `<div class="baca-juga"></div>`. Pola *regex* yang dibentuk cukup menggunakan *tag* pembuka dan penutup serta atribut dari *tag* ini.

3. Sisipan gambar

Sisipan gambar pada halaman ini diletakkan pada *tag* ``. Pola *regex*

yang dibentuk cukup disusun menggunakan *tag* ini.

3. Tribunnews

Hasil analisis struktur halaman ini menunjukkan bahwa pola penulisan isi berita yang dapat dijadikan acuan pembuatan pola *regex* adalah `<div class="teks-article mb20">Isi Berita </div>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<div>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak isi berita yang ditampilkan.

Untuk membersihkan berita dari konten non-berita, dibutuhkan pola *regex* tersendiri untuk menemukan kemudian menghapusnya. Berikut adalah proses pembentukan pola *regex* tersebut.

1. Sisipan iklan

Sisipan iklan pertama diletakkan menggunakan *tag* `<div id="div-Inside-MediumRectangle" style="teks-align: center; margin: auto;" data-google-query-id="COyBwKPS9NsCFQcKjgodX7Q01A">Isi Iklan</div>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup serta beberapa atribut dari *tag* ini. Sisipan iklan kedua ditulis menggunakan *tag* `<div class='adspruce-bannerspot'></div>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup serta beberapa atribut dari *tag* ini.

2. Sisipan tautan lain

Sisipan tautan pertama diletakkan di antara *tag* `<p class="baca"> Baca :judul berita</p>`. Pola *regex* yang dibentuk cukup disusun menggunakan kata "Baca" ditambahkan dengan *tag* pembuka dan penutup dari *tag* ini. Sisipan tautan kedua ditulis menggunakan *tag* `<a>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dari *tag* ini. Sisipan tautan ketiga ditulis menggunakan *tag* `<u1></u1>`. Pola *regex* yang dibentuk cukup menggunakan *tag* pembuka dan penutup serta atribut dari *tag* ini.

3. Sisipan gambar

Sisipan gambar pada halaman ini diletakkan pada *tag* `<figure>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup *tag* ini.

4. Sisipan tag ``

Sisipan gambar pada halaman ini diletakkan pada *tag* ``. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup serta atribut dari *tag* ini.

5. Sisipan script

Sisipan gambar pada halaman ini diletakkan pada *tag* `<script src="Sumber Script" async="true"></script>`. Pola *regex* yang dibentuk cukup disusun menggunakan *tag* pembuka dan penutup serta atribut dari *tag* ini.

4.2.4. Pola *Regex* Penulis Berita

1. Detik

Struktur html detik.com memiliki dua jenis pola peletakan penulis berita, yaitu penulis berita video dan penulis berita narasi. Berdasarkan hasil analisis struktur halaman dari dua jenis berita tersebut, berikut adalah proses pembentukan pola *regex* yang dibutuhkan.

➤ Berita Video

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan penulis berita pada halaman berita video yang dapat dijadikan acuan adalah `Nama Penulis/Sumber Berita`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* `` pembuka setelahnya digabungkan dengan *tag* pembuka dan penutup dari *tag* `` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

➤ Berita Narasi / Tekstual

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan penulis berita pada halaman berita narasi yang dapat dijadikan acuan adalah `<div class="author">Nama Penulis</div>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<div>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

2. Liputan 6

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan penulis berita pada halaman berita video dan

narasi yang dapat dijadikan acuan adalah `Nama Penulis/Sumber`. Berdasarkan pola penulisan tersebut, maka dibutuhkan dua pola *regex* untuk melakukan ekstraksi.

Pola *regex* yang pertama berfungsi untuk mencari isi dari *tag* ``. Isi yang ditemukan di dalam *tag* ini ada dua macam, yaitu nama penulis atau tautan menuju blog penulis. Dibutuhkan pola kedua untuk menghapus tautan blog penulis.

Pola *regex* yang kedua berfungsi untuk mencari dan menghapus *tag* `<a>` yang ada pada hasil ekstraksi pola pertama (yang berada di dalam *tag* ``). Pola ini dibentuk cukup dengan *tag* pembuka dan penutup dari *tag* `<a>` yang digunakan.

3. Tribunnews

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan penulis berita pada halaman berita video dan narasi yang dapat dijadikan acuan adalah `<div class="f12 grey mb15"> Nama Penulis/Sumber</div>`. Berdasarkan pola penulisan tersebut, maka dibutuhkan dua pola *regex* untuk melakukan ekstraksi.

Pola *regex* yang pertama berfungsi untuk mencari isi dari *tag* `<div class="f12 grey mb15"> Nama Penulis/Sumber</div>`. Isi yang ditemukan di dalam *tag* ini ada dua macam, yaitu nama penulis atau tautan menuju sumber berita. Jika hanya nama penulis tanpa tautan sumber berita, maka pola yang dibentuk cukup menggunakan *tag* pembuka dan penutup serta atribut dari *tag* ini. Sedangkan jika di dalamnya mengandung tautan sumber berita maka dibutuhkan pola kedua untuk menghapus tautan sumber berita.

Pola *regex* yang kedua berfungsi untuk mencari dan menghapus *tag* `<a>` yang ada pada hasil ekstraksi pola pertama (yang berada di dalam *tag* `<div class="f12 grey mb15"> Nama Penulis/Sumber</div>`). Pola ini dibentuk cukup dengan *tag* pembuka dari *tag* `<a>` yang digunakan.

4.2.5. Pola *Regex* Tanggal Publikasi Berita

1. Detik

Struktur html detik.com memiliki dua jenis pola peletakan tanggal publikasi berita, yaitu untuk berita video dan berita narasi. Berdasarkan hasil analisis struktur halaman dari dua jenis berita tersebut, berikut adalah proses pembentukan pola *regex* yang dibutuhkan.

➤ Berita Video

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tanggal publikasi berita pada halaman berita video yang dapat dijadikan acuan adalah `20DETIK Tanggal Publikasi`. Berdasarkan pola penulisan tersebut, maka dengan menggabungkan kata `20DETIK`, *tag* `` penutup, kemudian digabungkan dengan *tag* pembuka dan penutup dari *tag* `` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak tanggal publikasi berita yang ditampilkan.

➤ Berita Narasi / Tekstual

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tanggal publikasi berita pada halaman berita narasi yang dapat dijadikan acuan adalah `<div class="date">Tanggal Publikasi Berita</div>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<div>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak tanggal publikasi berita yang ditampilkan.

2. Liputan 6

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tanggal publikasi berita pada halaman berita video dan narasi yang dapat dijadikan acuan adalah `Tanggal Publikasi Berita`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

3. Tribunnews

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tanggal publikasi berita pada halaman berita video dan narasi yang dapat dijadikan acuan adalah `<time class="grey f13 dip">Tanggal Publikasi Berita</time>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup dari *tag* `<time>` beserta atribut yang dipakai sudah cukup untuk menyusun pola *regex* untuk mengekstrak judul berita yang ditampilkan.

4.2.6. Pola *Regex* Tautan Berita Berupa Video

1. Detik

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tautan video berita pada halaman berita video yang dapat dijadikan acuan adalah `<iframe src="tautan video" frameborder="0" width="420" scrolling="no" allowfullscreen="true" height="236"></iframe>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan penutup beserta atribut dari *tag* `<iframe>` sudah cukup untuk menyusun pola *regex* untuk mengekstrak tanggal publikasi berita yang ditampilkan.

2. Liputan 6

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tautan video berita pada halaman berita video yang dapat dijadikan acuan adalah `<iframe class="vidio-embed article-video-gallery-item__vidio-embed" src="Tautan Video" width="100%" height="205" frameborder="0" scrolling allowfullscreen data-component-name="ui:frame:vidio"></iframe>`. Berdasarkan pola penulisan tersebut, maka dengan menggunakan *tag* pembuka dan beberapa atribut dari *tag* `<iframe>` sudah cukup untuk menyusun pola *regex* untuk mengekstrak tanggal publikasi berita yang ditampilkan.

3. Tribunnews

Pada tahap analisis struktur halaman yang dilakukan sebelumnya, ditemukan bahwa pola penulisan tautan video berita pada halaman berita video yang dapat dijadikan acuan adalah `<div class="video" id="frame-resize"><iframe width="300" height="169" src="Tautan Video" frameborder="0" allowfullscreen=""`

style="width: 100%; height: 202.8px;"></iframe>. Berdasarkan pola penulisan tersebut, maka dengan menggabungkan *tag* <div class="video" id="frame-resize"> dan *tag* pembuka serta beberapa atribut dari *tag* <iframe> sudah cukup untuk menyusun pola *regex* untuk mengekstrak tanggal publikasi berita yang ditampilkan

4.3. Analisa Kebutuhan Aplikasi

Tahapan analisis ini dilakukan mengacu kepada tujuan pengembangan aplikasi. Hasil analisis ini berupa kebutuhan fungsional dan non-fungsional aplikasi berdasarkan fungsionalitas pada batasan masalah.

Tahapan analisis aplikasi dilakukan secara terpisah antara analisis kebutuhan untuk aplikasi *scraper* dan *monitoring*. Analisis kebutuhan pada *scraper* dilakukan dengan melakukan analisis terhadap konten berita apa saja yang diambil, pola dan navigasi masing-masing halaman situs untuk dijadikan *scraping template*, dan bagaimana cara aplikasi untuk mengotomatisasi navigasi dari masing-masing halaman tersebut. Analisis kebutuhan aplikasi *monitoring* dilakukan dengan menganalisis fitur-fitur apa yang dibutuhkan untuk memenuhi fungsionalitas aplikasi sesuai dengan batasan masalah.

4.3.1. Kebutuhan Fungsional

1. *Scraper*

- a. Aplikasi dapat mengambil dan menyimpan konten berita berupa judul, penulis/sumber, tanggal publikasi, tautan, tautan video (untuk berita video), isi berita, isi berita beserta *tag* html yang digunakan kedalam basis data.
- b. Aplikasi dapat memisahkan antara isi berita dengan sisipan gambar atau tautan berita lain yang disisipkan ke dalam konten berita.
- c. Aplikasi dapat membedakan jenis berita yang di ekstraksi berupa video atau teks
- d. Aplikasi dapat melakukan *scraping* terhadap berita baru pada masing-masing situs setiap 5 menit sekali.

2. *Monitoring*

- a. Aplikasi dapat membedakan akses dari 3 aktor (tamu, anggota, dan admin).
- b. Aplikasi dapat menangani pendaftaran tamu menjadi anggota.

- c. Aplikasi dapat menampilkan statistik hasil scraping dalam bentuk grafik batang.
- d. Aplikasi dapat menampilkan informasi berupa nama situs, alamat, *ranking*, *daily pageviews*, *traffic from search*, *daily time*, dan *site tautaning* masing-masing situs berita berdasarkan data dari www.alex.com.
- e. Aplikasi hanya dapat menampilkan daftar dan detail berita hasil *scraping* untuk level anggota.
- f. Aplikasi hanya dapat menampilkan dan memberikan akses mengubah profil (password dan foto profil akun) untuk level anggota.
- g. Aplikasi dapat menampilkan API untuk level anggota.
- h. Aplikasi hanya dapat menampilkan daftar dan detail pendaftaran anggota baru untuk akses admin.
- i. Aplikasi hanya dapat memberikan akses persetujuan / penolakan keanggotaan untuk level admin.
- j. Aplikasi hanya dapat memberikan akses penonaktifan / pengaktifan kenggotaan untuk level admin
- k. Aplikasi dapat membuat API baru untuk anggota pada saat admin menyetujui kenggotaan.
- l. Aplikasi hanya dapat menampilkan daftar anggota untuk level admin.
- m. Aplikasi hanya dapat menampilkan, memberikan akses menambah, merubah, dan menghapus jenis berita untuk level admin
- n. Aplikasi hanya dapat menampilkan, memberikan akses menambah, merubah, dan menghapus pola *regex* untuk level admin
- o. Aplikasi hanya dapat menampilkan, memberikan akses menambah, merubah, dan menghapus pola khusus untuk level admin
- p. Aplikasi hanya dapat menampilkan, memberikan akses menambah, merubah, dan menghapus filter html berita untuk level admin

4.3.2. Kebutuhan Non-fungsional

1. *Scraper*

- a. Aplikasi dikembangkan dalam bahasa *website c#*.
- b. Aplikasi berupa *Command Line Interface (CLI)*.

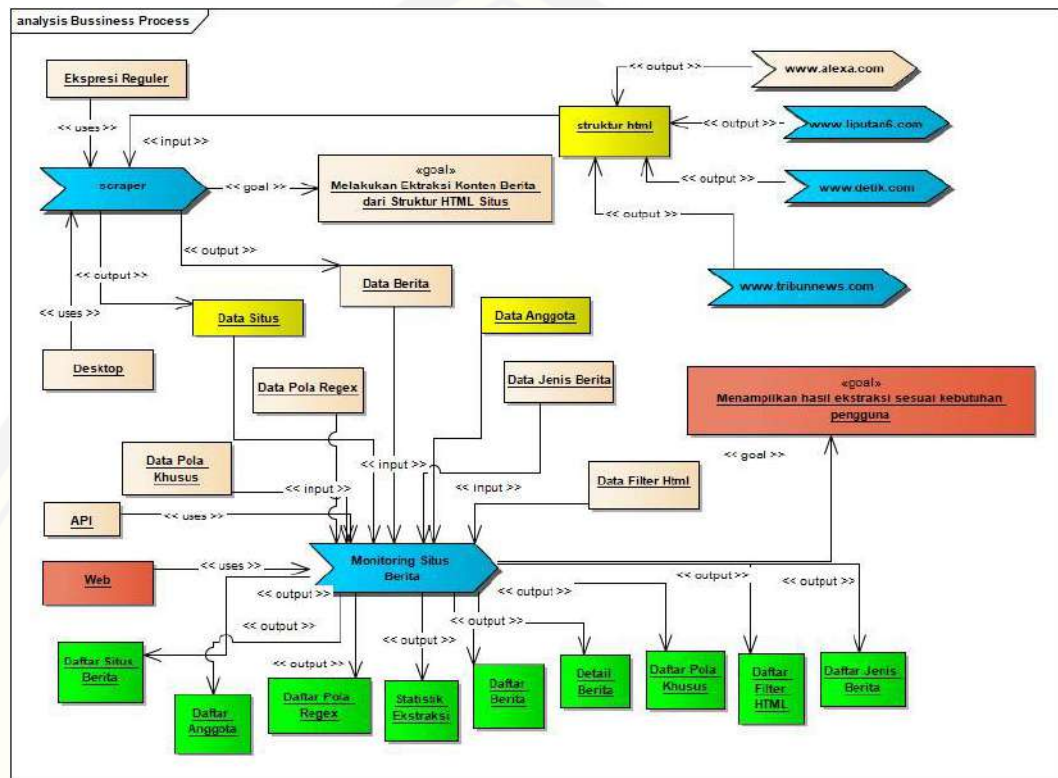
- c. Aplikasi dapat menampilkan log aktivitas pada saat proses *scarping* dilakukan
2. *Monitoring*
 - a. Aplikasi dikembangkan menggunakan bahasa *website c#* dan *framework ASP.Net*.
 - b. Aplikasi memiliki tampilan antar muka yang menarik dan mudah bagi pengguna.
 - c. Aplikasi dapat dijalankan oleh beberapa *software web browser* di antaranya *Microsft Edge, Google Chrome, dan Mozilla Firefox*.
 - d. Basis data aplikasi dibangun dengan *postgresql*.

4.4. Desain Aplikasi

Tahapan yang dilakukan setelah melakukan analisis kebutuhan aplikasi adalah tahap perencanaan aplikasi yang digambarkan dalam desain aplikasi. Desain aplikasi digambarkan melalui *usecase skenario, activity diagram, sequence diagram, class diagram* dan *entity relationship diagram*.

4.4.1. Bussiness Process

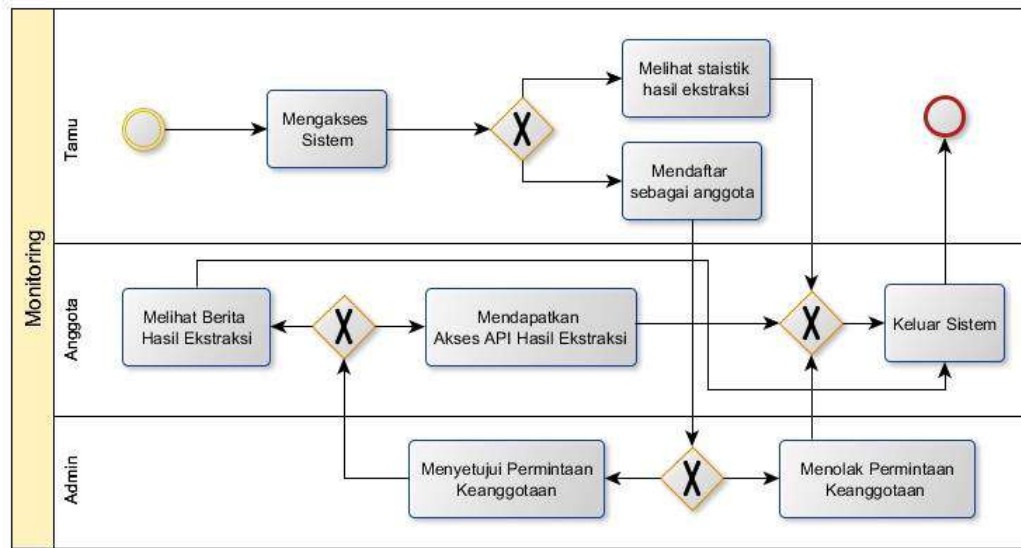
Business Process adalah suatu kumpulan aktivitas yang terstruktur untuk mencapai suatu tujuan tertentu atau untuk menghasilkan sebuah produk. Ada beberapa komponen di dalamnya, meliputi data yang menjadi masukan (*input*), data masukan yang kemudian diolah menjadi data keluaran (*output*), media yang digunakan (*uses*), tujuan yang ingin dicapai (*goal*). *Business Process* aplikasi ini dapat pada Gambar 4.24.



Gambar 4. 24 *Bussiness Process*

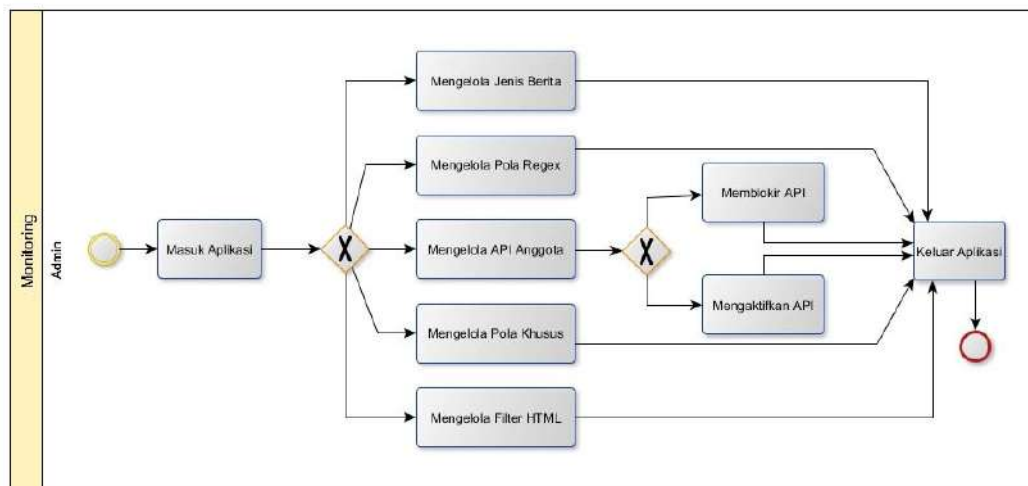
Gambar 4.24 menggambarkan ada dua sub aplikasi yang dibangun yaitu sub aplikasi *scraper* dan sub aplikasi *monitoring*. *Scraper* dibangun menggunakan *uses regex* dan berbasis *desktop* serta membutuhkan satu masukan utama berupa struktur html dan dua keluaran berupa data situs berita dan data berita. *Monitoring* dibangun dengan menggunakan *uses API* dan berbasis web serta membutuhkan tujuh masukan berupa data berita, data situs berita, data anggota, data pola *regex*, data pola khusus, data jenis berita dan data filter html. Keluaran yang dihasilkan berupa daftar berita, grafik ekstraksi, daftar situs berita, daftar anggota, detail anggota, detail berita, daftar pola *regex*, daftar pola khusus, daftar filter html dan daftar jenis berita.

Pada sub aplikasi monitoring, terdapat dua proses utama yang ditangani, yaitu proses pengelolaan anggota dan akses API yang dimiliki, serta pengelolaan pola *regex* yang digunakan oleh *scraper*. Dua proses utama tersebut dapat dilihat pada Gambar 4.25 dan Gambar 4.26.



Gambar 4. 25 BPMN pengelolaan anggota

Gambar 4.25 menggambarkan proses utama dalam mengelola anggota. Pengelolaan anggota dimulai dengan aktor tamu mendaftar kemudian disetujui atau ditolak admin, jika disetujui maka pendaftar akan menjadi anggota dan mendapatkan akses API yang dimiliki aplikasi.

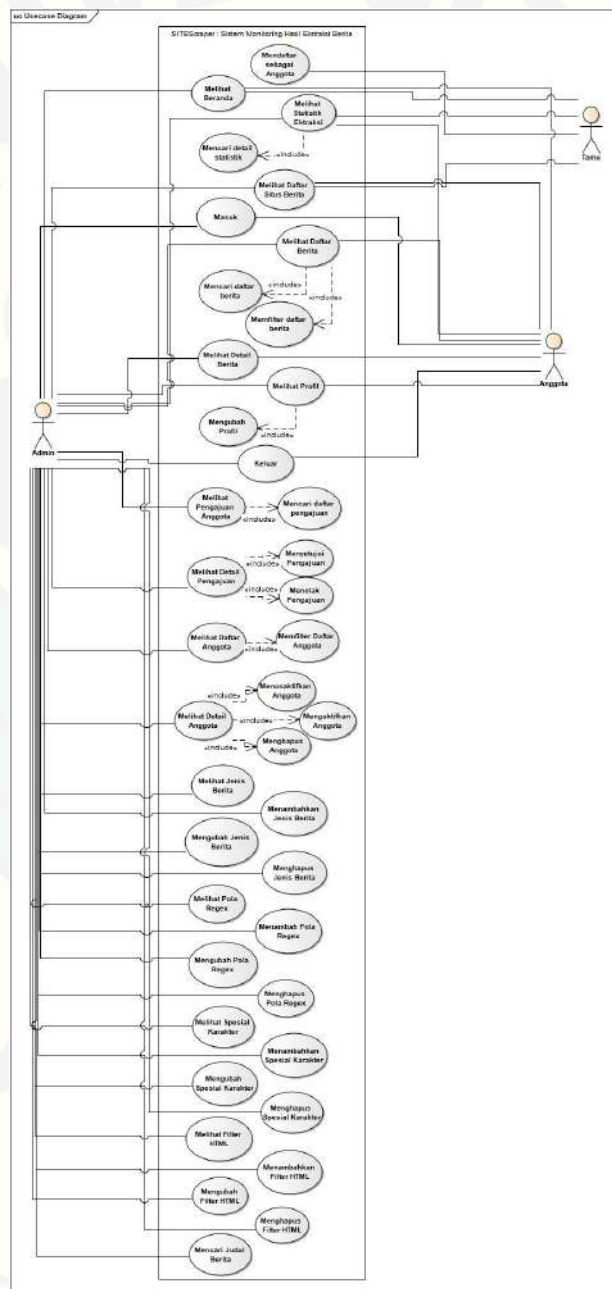


Gambar 4. 26 BPMN pengelolaan pola *regex*

Gambar 4.26 menggambarkan proses utama pada pengelolaan pola *regex* yang digunakan oleh sub aplikasi *scraper*. Pada proses ini aktor yang berperan adalah admin. Admin dapat mengelola jenis berita yang akan diambil, pola regex yang digunakan, pola khusus yang dibutuhkan serta filter html yang dipakai.

4.4.2. Usecase Diagram

Usecase diagram merupakan pemodelan yang dibuat untuk dapat menggambarkan interaksi antara aktor dengan Sistem Informasi yang akan dibangun. Melalui *usecase* diagram dapat diketahui interaksi yang dapat dilakukan aktor terhadap aplikasi sesuai dengan hak akses yang dimiliki oleh masing-masing aktor atau pengguna. *Usecase diagram* aplikasi ini dapat dilihat pada gambar 4.27.



Gambar 4. 27 Usecase Diagram

Penjelasan tentang definisi aktor dan definisi *usecase* dalam *usecase* diagram pada Gambar 4.27 dapat dilihat pada tabel 4.1.

Tabel 4. 1 Deskripsi Aktor

Aktor	Deskripsi
Tamu	Aktor yang hanya bisa mengakses statistik hasil ekstraksi dan informasi <i>website</i> . Aktor ini juga memiliki akses untuk mendaftar sebagai anggota untuk mendapatkan API.
Anggota	Aktor yang mempunyai akses ke semua fitur pada level tamu, ditambah dengan akses untuk membaca berita hasil <i>scraping</i> dan mendapatkan akses API aplikasi.
Admin	Aktor yang bertanggungjawab terhadap pengelolaan keanggotaan, pengelolaan jenis berita, dan pengelolaan pola ekspresi reguler yang digunakan oleh <i>scraper</i> . Aktor ini juga memiliki akses ke semua fitur pada level tamu dan anggota.

Definisi *Usecase* merupakan penjelasan dari masing-masing *Usecase* atau fitur-fitur dari aplikasi yang akan dibangun. Definisi masing-masing *Usecase* terdapat pada Tabel 4.2.

Tabel 4. 2 Deskripsi Usecase

Usecase	Deskripsi
Mendaftar sebagai anggota	<i>Usecase</i> ini menggambarkan proses aktor tamu melengkapi formulir pendaftaran sebagai anggota.
Melihat beranda	<i>Usecase</i> ini menggambarkan proses semua aktor untuk melihat rangkuman hasil ekstraksi dalam bentuk grafik
Melihat statistik ekstraksi	<i>Usecase</i> ini menggambarkan proses aktor melihat detail hasil ekstraksi terhadap masing-masing situs berita.
Melihat daftar situs berita	<i>Usecase</i> ini menggambarkan proses aktor melihat daftar situs berita.
Masuk	<i>Usecase</i> ini menggambarkan aktor anggota/admin masuk ke dalam akses level masing-masing.
Melihat daftar berita	<i>Usecase</i> ini menggambarkan proses aktor melihat daftar berita yang berhasil di ekstraksi.
Melihat detail berita	<i>Usecase</i> ini menggambarkan proses aktor melihat detail berita yang berhasil di ekstraksi.
Melihat profil	<i>Usecase</i> ini menggambarkan proses aktor melihat profil masing-masing.
Keluar	<i>Usecase</i> ini menggambarkan proses aktor (admin/anggota) keluar dari akses level masing-masing.
Melihat pengajuan anggota	<i>Usecase</i> ini menggambarkan proses aktor melihat daftar pendaftar yang masuk dan mendaftar sebagai anggota.

Usecase	Deskripsi
Melihat detail pengajuan	<i>Usecase</i> ini menggambarkan proses aktor melihat detail pendaftar
Menyetujui pengajuan	<i>Usecase</i> ini menggambarkan proses aktor melakukan persetujuan terhadap pendaftar dan merubah levelnya menjadi anggota aktif.
Menolak pengajuan	<i>Usecase</i> ini menggambarkan proses aktor melakukan penolakan terhadap pendaftar dan menghapus data pendaftar dari dalam basis data.
Melihat daftar anggota	<i>Usecase</i> ini menggambarkan proses aktor melihat daftar anggota yang sudah terdaftar dan memilki API
Melihat detail anggota	<i>Usecase</i> ini menggambarkan proses aktor melihat detai anggota.
Menonaktifkan anggota	<i>Usecase</i> ini menggambarkan proses aktor menonaktifkan status keanggotaan sekaligus menonaktifkan akses API.
Mengaktifkan anggota	<i>Usecase</i> ini menggambarkan proses aktor mengaktifkan status keanggotaan sekaligus mengaktifkan akses API.
Menghapus anggota	<i>Usecase</i> ini menggambarkan proses aktor menghapus keanggotaan seorang anggota
Menambahkan jenis berita	<i>Usecase</i> ini menggambarkan proses aktor memasukkan jenis berita baru yang terdapat pada salah satu situs berita.
Mengubah jenis berita	<i>Usecase</i> ini menggambarkan proses aktor mengubah nama dari jenis berita yang telah dimasukkan sebelumnya.
Menghapus jenis berita	<i>Usecase</i> ini menggambarkan proses aktor untuk menghapus jenis berita yang telah dimasukkan sebelumnya.
Menambahkan pola ekspresi reguler	<i>Usecase</i> ini menggambarkan proses aktor untuk menambahkan pola ekspresi reguler baru untuk digunakan <i>scraper</i> dalam mengekstraksi situs berita.
Mengubah pola ekspresi reguler	<i>Usecase</i> ini menggambarkan proses aktor mengubah pola ekspresi reguler yang telah dimasukkan sebelumnya.
Menghapus pola ekspresi reguler	<i>Usecase</i> ini menggambarkan proses aktor menghapus pola ekspresi reguler yang telah dimasukkan sebelumnya.
Menambahkan pola khusus	<i>Usecase</i> ini menggambarkan proses aktor untuk menambahkan pola khusus baru untuk digunakan <i>scraper</i> dalam mengekstraksi situs berita.
Mengubah pola khusus	<i>Usecase</i> ini menggambarkan proses aktor mengubah pola khusus yang telah dimasukkan sebelumnya.
Menghapus pola khusus	<i>Usecase</i> ini menggambarkan proses aktor menghapus pola khusus yang telah dimasukkan sebelumnya.
Menambahkan filter html	<i>Usecase</i> ini menggambarkan proses aktor untuk menambahkan filter html baru untuk digunakan <i>scraper</i> dalam mengekstraksi situs

Usecase	Deskripsi
	berita.
Mengubah filter html	<i>Usecase</i> ini menggambarkan proses aktor mengubah filter html yang telah dimasukkan sebelumnya.
Menghapus filter html	<i>Usecase</i> ini menggambarkan proses aktor menghapus filter html yang telah dimasukkan sebelumnya.
Mencari judul berita	<i>Usecase</i> ini menggambarkan proses aktor melakukan pencarian berita berdasarkan judul berita yang berhasil diekstrak.

4.4.3. Usecase Skenario

Usecase skenario digunakan untuk menjelaskan alur aplikasi sesuai dengan yang ada pada *usecase diagram* seperti gambar 4.27.

1) Melakukan ekstraksi situs berita

Skenario *usecase* melakukan ekstraksi situs berita dilakukan oleh sub aplikasi *scraper*. Skenario ini menjelaskan bagaimana *scraper* melakukan ekstraksi konten berita pada masing-masing situs. Proses lengkapnya dapat dilihat pada tabel 4.3.

Tabel 4. 3 *Usecase* skenario melakukan ekstraksi situs berita

No Usecase	USC1
Nama Usecase	Melakukan ekstraksi situs berita
Aktor	<i>Task Scheduler</i>
Deskripsi Singkat	<i>Task Scheduler</i> menjalankan program setiap 5 menit untuk melakukan proses <i>scraping</i> data berita.
Prakondisi	<i>Task</i> pada <i>task scheduler</i> dalam status <i>ready</i>
Paskakondisi	Program <i>scraper</i> terbuka dan menampilkan hasil dari proses ekstraksi
Aktor	Aplikasi
Normal Flow	
1. <i>Run Scraper</i>	2. Memperbaharui data detik.com berdasarkan data pada situs www.alex.com dengan alamat halaman yang dituju adalah https://www.alex.com/siteinfo/detik.com Informasi yang diekstrak berupa : <ul style="list-style-type: none"> - <i>Site Rank</i> dengan pola <i>regex</i> “<strong class=\"metrics-data align-vmiddle\">(.*?)” - <i>Daily Time</i> dengan pola <i>regex</i> “<h4 class=\"metrics-title\">Daily Time on Site</h4><div><strong class=\"metrics-data align-vmiddle\">(.*?)” - <i>Daily Pageviews</i> dengan pola <i>regex</i> berupa

	<pre> <h4 class=\"metrics-title\">Daily Pageviews per Visitor</h4><div><strong class=\"metrics-data align- vmiddle\">(.*?)” - Site Traffic dengan pola regex berupa <h4 class=\"metrics-title\">Search Visits</h4><div><strong class=\"metrics- data align-vmiddle\">(.*?)” - Site Tautaning dengan pola regex berupa <h5 class=\"font-1 box1-1\">Total Sites Tautaning In</h5>(.*?)” 3. Menampilkan pesan “detik has been updated” 4. Melakukan proses <i>scraping</i> tautan berita pada situs detik.com dengan <i>user agent</i> berupa <i>mobile</i> pada halaman indeks berita narasi dengan alamat halaman indeks pada basis data + “/berita” 5. Menampilkan pesan “get html page from “+alamat halaman indeks+” ~SUCCESS~” 6. Melakukan proses <i>scraping</i> data berita pada masing-masing halaman dari tautan yang sudah di dapatkan dengan pola pengambilan konten berita dengan langkah : - Mengambil judul berita sesuai dengan pola ekspresi reguler pada basis data. - Mengambil tanggal publikasi berita sesuai dengan pola ekspresi reguler pada basis data. - Mengambil penulis berita sesuai dengan pola ekspresi reguler pada basis data. - Mengambil isi berita sesuai dengan pola ekspresi reguler pada basis data. 7. Menampilkan hasil ekstraksi berupa: a. Judul b. Penulis c. Tanggal d. Tautan e. Isi Berita f. Tautan Video (Jika ada) 8. Melakukan langkah 4-7 dengan alamat halaman indeks yang di rubah menjadi alamat indeks pada basis data + “/detiktv”. 9. Melakukan update data alexa untuk situs liputan6.com seperti pada langkah 1, dengan alamat halaman alexa https://www.alex.com/siteinfo/liputan6.com 10. Melakukan proses <i>scraping</i> mulai dari langkah 4-7 dengan alamat halaman indeks liputan6 pada basis data. 11. Melakukan update data alexa untuk situs tribunnews.com seperti pada langkah 1, dengan alamat halaman alexa https://www.alex.com/siteinfo/tribunnews.com 12. Melakukan proses <i>scraping</i> mulai dari langkah 4-7 dengan alamat halaman indeks tribunnews pada basis data. 13. Menampilkan pesan “All site has been scraped” </pre> <p>Alternatif flow : Gagal mengakses tautan</p>
1. <i>Run Scraper</i>	
	2. Menampilkan pesan error dari <i>Exception.Web</i>

2) Melihat Beranda

Skenario *Usecase* melihat beranda dilakukan oleh semua aktor. Skenario *Usecase* melihat beranda menjelaskan tentang bagaimana aplikasi dapat menampilkan ringkasan statistik hasil *scraping* dalam bentuk grafik. Proses lengkapnya dapat dilihat pada Lampiran A.

3) Mendaftar Sebagai Anggota

Skenario *Usecase* mendaftar sebagai anggota dilakukan oleh semua aktor. Skenario *Usecase* mendaftar sebagai anggota menjelaskan bagaimana aplikasi menampilkan formulir pendaftaran dan menangani pendaftaran aktor sesuai dengan data yang dimasukkan pada formulir pendaftaran. Proses lengkapnya dapat dilihat pada Lampiran A.

4) Melihat Statistik Ekstraksi

Skenario *Usecase* melihat statistik ekstraksi dapat dilakukan oleh semua aktor. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menampilkan detail statistik hasil ekstraksi dalam bentuk grafik harian, bulanan dan tahunan dari masing-masing situs berita. Proses lengkapnya dapat dilihat pada Lampiran A.

5) Mencari Detail Statistik

Skenario mencari detail statistik dapat dilakukan oleh semua aktor, skenario ini hanya berjalan ketika skenario melihat statistik ekstraksi sudah dijalankan. Proses lengkapnya dapat dilihat pada Lampiran A.

6) Melihat Daftar Situs Berita

Skenario *Usecase* melihat daftar situs berita dapat dilakukan oleh semua aktor. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor pada proses menampilkan daftar situs berita dan informasi apa saja yang ditampilkan oleh aplikasi. Proses lengkapnya dapat dilihat pada Lampiran A.

7) Masuk Aplikasi

Skenario masuk aplikasi dapat dilakukan oleh anggota dan admin. Skenario ini menjelaskan bagaimana aplikasi dapat membedakan hak akses antara anggota dan admin. Proses lengkapnya dapat dilihat pada Lampiran A.

8) Melihat Daftar Berita

Skenario melihat daftar berita dapat dilakukan oleh anggota dan admin.

Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan daftar berita yang berhasil di ekstraksi dari situs berita. Proses lengkapnya dapat dilihat pada Lampiran A.

9) Mencari Daftar Berita

Skenario mencari daftar berita menggambarkan bagaimana aktor berinteraksi dengan aplikasi pada saat melakukan proses pencarian berita. Skenario ini hanya bisa berjalan jika skenario melihat daftar berita sudah dijalankan. Proses lengkapnya dapat dilihat pada Lampiran A.

10) Memfilter Daftar Berita

Skenario memfilter daftar berita menggambarkan bagaimana aktor berinteraksi dengan aplikasi pada saat melakukan proses filter berita. Skenario ini hanya bisa berjalan jika skenario melihat daftar berita sudah dijalankan. Proses lengkapnya dapat dilihat pada Lampiran A.

11) Melihat Detail Berita

Skenario melihat detail berita dapat dilakukan oleh anggota dan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan detail dari berita yang telah berhasil di ekstraksi. Proses lengkapnya dapat dilihat pada Lampiran A.

12) Melihat Profil

Skenario melihat profil dapat dilakukan oleh anggota dan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan profil dari aktor. Proses lengkapnya dapat dilihat pada Lampiran A.

13) Mengubah Profil

Skenario mengubah profil dapat dilakukan oleh anggota dan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk proses perubahan profil berupa foto profil dan kata sandi aktor. Proses lengkapnya dapat dilihat pada Lampiran A.

14) Keluar Aplikasi

Skenario keluar aplikasi dapat dilakukan oleh anggota dan admin yang telah masuk ke dalam aplikasi. Skenario ini menjelaskan bagaimana aplikasi menghapus akses aktor ketika aktor telah keluar dari aplikasi. Proses lengkapnya dapat dilihat

pada Lampiran A.

15) Melihat Daftar Pengajuan Anggota

Skenario melihat daftar pengajuan anggota dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi menampilkan daftar pendaftar yang mengajukan keanggotaan. Proses lengkapnya dapat dilihat pada Lampiran A.

16) Mencari Daftar Pengajuan

Skenario mencari daftar pengajuan menggambarkan bagaimana aktor berinteraksi dengan aplikasi pada saat melakukan proses pencarian daftar pengajuan. Skenario ini hanya bisa berjalan jika skenario melihat daftar pengajuan anggota sudah dijalankan. Proses lengkapnya dapat dilihat pada Lampiran A.

17) Melihat Detail Pengajuan Anggota

Skenario melihat detail pengajuan anggota dapat dilakukan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menampilkan data diri pendaftar secara detail berdasarkan data pendaftaran. Proses lengkapnya dapat dilihat pada Lampiran A.

18) Menyetujui Pengajuan Anggota

Skenario menyetujui pengajuan anggota dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam mengubah status pendaftar menjadi anggota dan mendapatkan akses API aplikasi. Proses lengkapnya dapat dilihat pada Lampiran A.

19) Menolak Pengajuan Anggota

Skenario menolak pengajuan anggota dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menghapus data pendaftar dari basis data dan memberikan pemberitahuan terhadap pendaftar informasi penolakan pengajuan. Proses lengkapnya dapat dilihat pada Lampiran A.

20) Melihat Daftar Anggota

Skenario melihat daftar anggota dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi menampilkan daftar anggota yang tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

21) Memfilter Daftar Anggota

Skenario memfilter daftar anggota menggambarkan bagaimana aktor

berinteraksi dengan aplikasi pada saat melakukan proses filter anggota. Skenario ini hanya bisa berjalan jika skenario melihat daftar anggota sudah dijalankan. Proses lengkapnya dapat dilihat pada Lampiran A.

22) Melihat Detail Anggota

Skenario melihat detail anggota dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menampilkan detail anggota beserta log aktivitas API dari anggota tersebut. Proses lengkapnya dapat dilihat pada Lampiran A.

23) Menonaktifkan Keanggotaan

Skenario menonaktifkan keanggotaan dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mengubah status keanggotaan anggota menjadi tidak aktor sekaligus menonaktifkan akses API dari anggota tersebut. Proses lengkapnya dapat dilihat pada Lampiran A.

24) Mengaktifkan Keanggotaan

Skenario mengaktifkan keanggotaan dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mengubah status keanggotaan anggota menjadi aktor sekaligus mengaktifkan akses API dari anggota tersebut. Proses lengkapnya dapat dilihat pada Lampiran A.

25) Menghapus Keanggotaan

Skenario menghapus keanggotaan dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menghapus data anggota dari basis data dan memberikan pemberitahuan terhadap pendaftar informasi penghapusan keanggotaan kepada anggota. Proses lengkapnya dapat dilihat pada Lampiran A.

26) Melihat Daftar Jenis Berita

Skenario melihat daftar jenis berita dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan daftar jenis berita yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

27) Menambahkan Jenis Berita Baru

Skenario menambahkan jenis berita baru dapat dilakukan oleh admin.

Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menambahkan jenis berita baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

28) Mengubah Jenis Berita

Skenario mengubah jenis berita dapat dilakukan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mengubah data jenis berita yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

29) Menghapus Jenis Berita

Skenario menghapus jenis berita dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menghapus data jenis berita yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

30) Melihat Daftar Pola Ekspresi Reguler

Skenario melihat daftar pola ekspresi reguler dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan daftar pola ekspresi reguler yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

31) Menambahkan Pola Ekspresi Reguler

Skenario menambahkan pola ekspresi reguler baru dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menambahkan pola ekspresi reguler baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

32) Mengubah Pola Ekspresi Reguler

Skenario mengubah pola ekspresi reguler dapat dilakukan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mengubah data pola ekspresi reguler yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

33) Menghapus Pola Ekspresi Reguler

Skenario menghapus pola ekspresi reguler dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk

menghapus data pola ekspresi reguler yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

34) Melihat Daftar Pola Khusus

Skenario melihat daftar Pola Khusus dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan daftar Pola Khusus yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

35) Menambahkan Pola Khusus Baru

Skenario menambahkan Pola Khusus baru dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menambahkan Pola Khusus baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

36) Mengubah Pola Khusus

Skenario mengubah Pola Khusus dapat dilakukan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mengubah data Pola Khusus yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

37) Menghapus Pola Khusus

Skenario menghapus Pola Khusus dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menghapus data Pola Khusus yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

38) Melihat Daftar Filter HTML

Skenario melihat daftar filter html dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor dalam menampilkan daftar filter html yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

39) Menambahkan Filter HTML Baru

Skenario menambahkan filter html baru dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menambahkan filter html baru dan menyimpannya ke dalam basis data. Proses

lengkapanya dapat dilihat pada Lampiran A.

40) Mengubah Filter HTML

Skenario mengubah filter html dapat dilakukan admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mengubah data filter html yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

41) Menghapus Filter HTML

Skenario menghapus filter html dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk menghapus data filter html yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

42) Mencari Judul Berita

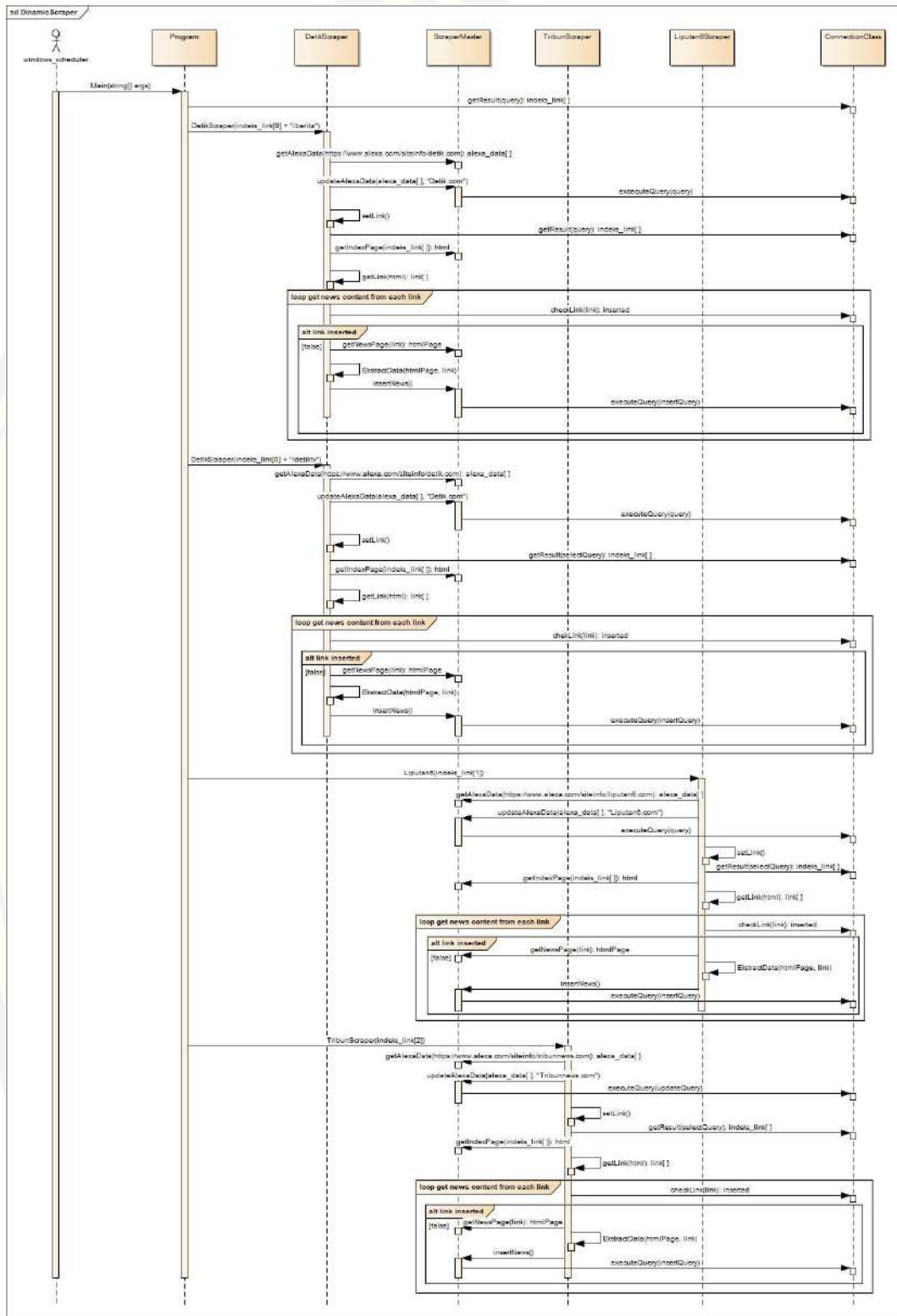
Skenario mencari judul berita dapat dilakukan oleh admin. Skenario ini menjelaskan bagaimana aplikasi berinteraksi dengan aktor untuk mencari judul berita yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran A.

4.4.4. Sequence Diagram

Sequence diagram adalah diagram yang digunakan untuk menggambarkan interaksi yang terjadi antar objek di dalam aplikasi. *Sequence* disusun dalam beberapa *class* yang terdiri dari *view*, *model* dan *controller*.

1) Melakukan Ekstraksi Situs Berita

Penggambaran *sequence diagram* melakukan ekstraksi situs berita digunakan untuk menjelaskan fungsi atau *method* yang dibuat untuk melakukan proses ekstraksi berita. Proses lengkapnya dapat dilihat pada Gambar 4.28.



Gambar 4. 28 Sequence Diagram melakukan ekstraksi Situs Berita

2) Mendaftar Sebagai Anggota

Penggambaran *sequence diagram* mendaftar sebagai anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses pendaftaran anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

3) Melihat Beranda

Penggambaran *sequence diagram* melihat beranda digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat beranda. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

4) Melihat Statistik Ekstraksi

Penggambaran *sequence diagram* melihat statistik ekstraksi digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat hasil ekstraksi. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

5) Mencari Detail Statistik

Penggambaran *sequence diagram* mencari statistik ekstraksi digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mencari detail statistik. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

6) Melihat Daftar Situs Berita

Penggambaran *sequence diagram* melihat daftar situs berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar situs berita. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

7) Masuk Aplikasi

Penggambaran *sequence diagram* masuk aplikasi digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor masuk ke dalam aplikasi. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

8) Melihat Daftar Berita

Penggambaran *sequence diagram* melihat daftar berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar berita. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

9) Mencari Daftar Berita

Penggambaran *sequence diagram* mencari daftar berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mencari daftar berita. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

10) Memfilter Daftar Berita

Penggambaran *sequence diagram* memfilter daftar berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor memfilter daftar berita. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

11) Melihat Detail Berita

Penggambaran *sequence diagram* melihat detail berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat detail berita. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

12) Melihat Profil

Penggambaran *sequence diagram* melihat profil digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat profil. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

13) Mengubah Profil

Penggambaran *sequence diagram* mengubah profil digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mengubah profil pengguna. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

14) Keluar Aplikasi

Penggambaran *sequence diagram* keluar aplikasi digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor keluar dari dalam aplikasi. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

15) Melihat Daftar Pengajuan Anggota

Penggambaran *sequence diagram* melihat daftar pengajuan anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar pengajuan anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

16) Mencari Daftar Pengajuan Anggota

Penggambaran *sequence diagram* mencari daftar pengajuan anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mencari daftar pengajuan anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

17) Melihat Detail Pengajuan Anggota

Penggambaran *sequence diagram* melihat detail pengajuan anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat detail pengajuan anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

18) Menyetujui Pengajuan Anggota

Penggambaran *sequence diagram* menyetujui pengajuan anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menyetujui pengajuan anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

19) Menolak Pengajuan Anggota

Penggambaran *sequence diagram* menolak pengajuan anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menolak pengajuan anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

20) Melihat Daftar Anggota

Penggambaran *sequence diagram* melihat daftar anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

21) Memfilter Daftar Anggota

Penggambaran *sequence diagram* memfilter daftar anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor memfilter daftar anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

22) Melihat Detail Anggota

Penggambaran *sequence diagram* melihat detail anggota digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat detail anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

23) Menonaktifkan Keanggotaan

Penggambaran *sequence diagram* menonaktifkan keanggotaan digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menonaktifkan anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

24) Mengaktifkan Keanggotaan

Penggambaran *sequence diagram* mengaktifkan keanggotaan digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mengaktifkan akses API anggota. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

25) Menghapus Keanggotaan

Penggambaran *sequence diagram* menghapus keanggotaan digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menghapus anggota dari keanggotaan. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

26) Melihat Daftar Jenis Berita

Penggambaran *sequence diagram* melihat daftar jenis berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar jenis berita. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

27) Menambahkan Jenis Berita Baru

Penggambaran *sequence diagram* menambahkan jenis berita baru digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menambahkan jenis berita baru. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

28) Mengubah Data Jenis Berita

Penggambaran *sequence diagram* mengubah data jenis berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mengubah data jenis berita yang telah tersimpan sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

29) Menghapus Data Jenis Berita

Penggambaran *sequence diagram* menghapus data jenis berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menghapus jenis berita yang telah dimasukkan sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

30) Melihat Daftar Pola Ekspresi Reguler

Penggambaran *sequence diagram* melihat daftar pola ekspresi reguler digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor

melihat daftar pola *regex*. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

31) Menambahkan Pola Ekspresi reguler Baru

Penggambaran *sequence diagram* menambahkan pola ekspresi reguler baru digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menambahkan pola *regex* baru. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

32) Mengubah Pola Ekspresi Reguler

Penggambaran *sequence diagram* mengubah pola ekspresi reguler digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mengubah pola *regex* yang telah ada sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

33) Menghapus Pola Ekspresi Reguler

Penggambaran *sequence diagram* menghapus pola ekspresi reguler digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menghapus pola *regex* yang telah tersimpan sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

34) Melihat Daftar Pola Khusus

Penggambaran *sequence diagram* melihat daftar pola khusus digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar pola khusus. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

35) Menambahkan Pola Khusus Baru

Penggambaran *sequence diagram* menambahkan pola khusus baru digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menambahkan pola khusus baru. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

36) Mengubah Pola Khusus

Penggambaran *sequence diagram* mengubah pola khusus digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mengubah pola khusus yang telah ada sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

37) Menghapus Pola Khusus

Penggambaran *sequence diagram* menghapus pola khusus digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menghapus pola khusus yang telah tersimpan sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

38) Melihat Daftar Filter HTML

Penggambaran *sequence diagram* melihat daftar filter html digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor melihat daftar filter html. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

39) Menambahkan Filter HTML Baru

Penggambaran *sequence diagram* menambahkan filter html baru digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menambahkan filter html baru. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

40) Mengubah Filter HTML

Penggambaran *sequence diagram* mengubah filter html digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mengubah filter html yang telah ada sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

41) Menghapus Filter HTML

Penggambaran *sequence diagram* menghapus filter html digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor menghapus filter html yang telah tersimpan sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

42) Mencari Judul Berita

Penggambaran *sequence diagram* mencari judul berita digunakan untuk menjelaskan fungsi atau *method* yang digunakan pada proses aktor mencari judul berita yang telah tersimpan sebelumnya. Proses lengkapnya seperti yang ditunjukkan pada Lampiran B.

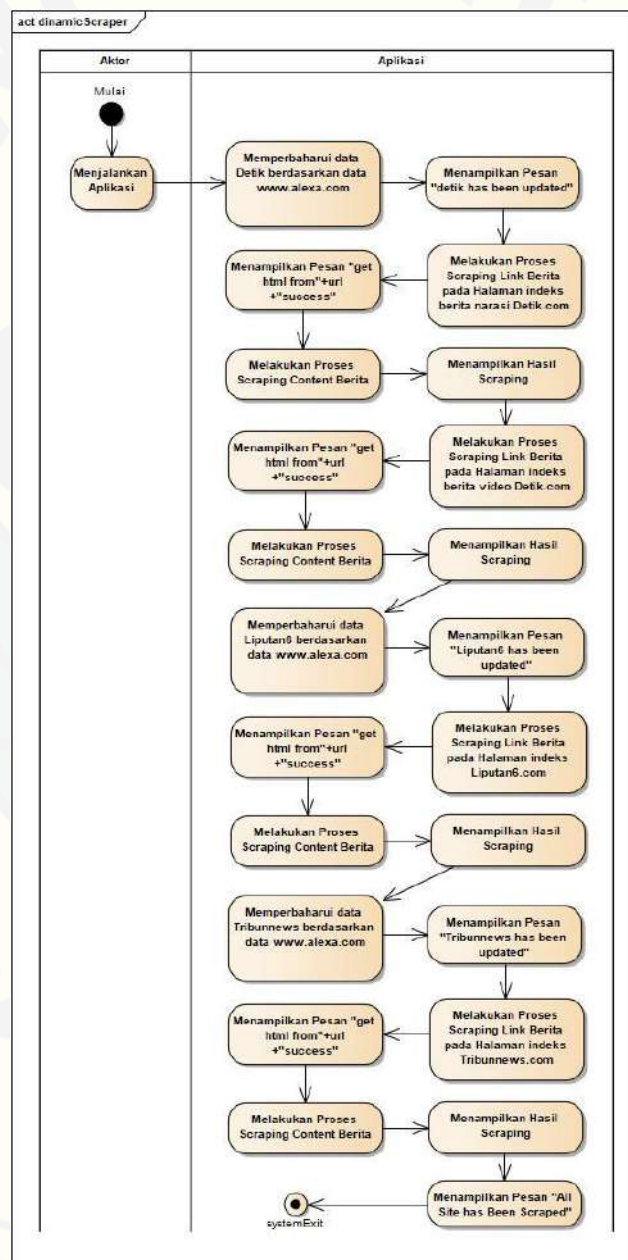
4.4.5. Activity Diagram

Activity diagram menggambarkan alur aktivitas pada aplikasi ekstraksi

konten pada situs berita dengan mengimplementasikan metode ekspresi reguler sebagai pencari pola.

1) Melakukan Ekstraksi Situs Berita

Activity Diagram melakukan ekstraksi situs berita dilakukan oleh aplikasi *scraper*. *Activity* ini menjelaskan apa saja aktivitas *scraper* dalam melakukan ekstraksi konten berita pada masing-masing situs. Proses lengkapnya dapat dilihat pada Gambar 4.29.



Gambar 4. 29 Activity Melakukan Ekstraksi Situs Berita

2) Melihat Beranda

Activity Diagram melihat beranda dilakukan oleh semua aktor. *Activity Diagram* melihat beranda menjelaskan tentang aktivitas apa saja pada saat aplikasi menampilkan ringkasan statistik hasil *scraping* dalam bentuk grafik. Proses lengkapnya dapat dilihat pada Lampiran C.

3) Mendaftar Sebagai Anggota

Activity Diagram mendaftar sebagai anggota dilakukan oleh semua aktor. *Activity Diagram* mendaftar sebagai anggota menjelaskan aktivitas aplikasi pada saat aplikasi menampilkan formulir pendaftaran dan menangani pendaftaran aktor sesuai dengan data yang dimasukkan pada formulir pendaftaran. Proses lengkapnya dapat dilihat pada Lampiran C.

4) Melihat Statistik Ekstraksi

Activity Diagram melihat statistik ekstraksi dapat dilakukan oleh semua aktor. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menampilkan detail statistik hasil ekstraksi dalam bentuk grafik harian, bulanan dan tahunan dari masing-masing situs berita. Proses lengkapnya dapat dilihat pada Lampiran C.

5) Mencari Statistik Ekstraksi

Activity Diagram mencari statistik ekstraksi dapat dilakukan oleh semua aktor. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menampilkan detail statistik hasil ekstraksi dalam bentuk grafik harian, bulanan dan tahunan dari masing-masing situs berita berdasarkan pencarian tanggal yang dilakukan pengguna. Proses lengkapnya dapat dilihat pada Lampiran C.

6) Melihat Daftar Situs Berita

Activity Diagram melihat daftar situs berita dapat dilakukan oleh semua aktor. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor pada proses menampilkan daftar situs berita dan informasi apa saja yang ditampilkan oleh aplikasi. Proses lengkapnya dapat dilihat pada Lampiran C.

7) Masuk Aplikasi

Activity Diagram masuk aplikasi dapat dilakukan oleh anggota dan admin.

Activity Diagram ini menjelaskan aktivitas aplikasi untuk membedakan hak akses antara anggota dan admin. Proses lengkapnya dapat dilihat pada Lampiran C.

8) Melihat Daftar Berita

Activity Diagram melihat daftar berita dapat dilakukan oleh anggota dan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar berita yang berhasil di ekstraksi dari situs berita. Proses lengkapnya dapat dilihat pada Lampiran C.

9) Mencari Daftar Berita

Activity Diagram mencari daftar berita dapat dilakukan oleh anggota dan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar berita yang berhasil di ekstraksi dari situs berita berdasarkan pencarian tanggal yang dilakukan oleh pengguna. Proses lengkapnya dapat dilihat pada Lampiran C.

10) Memfilter Daftar Berita

Activity Diagram melihat daftar berita dapat dilakukan oleh anggota dan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar berita yang berhasil di ekstraksi dari situs berita berdasarkan filter yang dipilih pengguna. Proses lengkapnya dapat dilihat pada Lampiran C.

11) Melihat Detail Berita

Activity Diagram melihat detail berita dapat dilakukan oleh anggota dan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan detail dari berita yang telah berhasil di ekstraksi. Proses lengkapnya dapat dilihat pada Lampiran C.

12) Melihat Profil

Activity Diagram melihat profil dapat dilakukan oleh anggota dan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan profil dari aktor. Proses lengkapnya dapat dilihat pada Lampiran C.

13) Mengubah Profil

Activity Diagram mengubah profil dapat dilakukan oleh anggota dan admin.

Activity Diagram ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk proses perubahan profil berupa foto profil dan kata sandi aktor. Proses lengkapnya dapat dilihat pada Lampiran C.

14) Keluar Aplikasi

Activity Diagram keluar aplikasi dapat dilakukan oleh anggota dan admin yang telah masuk ke dalam aplikasi. *Activity Diagram* ini menjelaskan aktivitas aplikasi untuk menghapus akses aktor ketika aktor telah keluar dari aplikasi. Proses lengkapnya dapat dilihat pada Lampiran C.

15) Melihat Daftar Pengajuan Anggota

Activity Diagram melihat daftar pengajuan anggota dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat menampilkan daftar pendaftar yang mengajukan keanggotaan. Proses lengkapnya dapat dilihat pada Lampiran C.

16) Mencari Daftar Pengajuan

Activity Diagram mencari daftar pengajuan menggambarkan aktivitas aktor berinteraksi dengan aplikasi pada saat melakukan proses pencarian daftar pengajuan. Skenario ini hanya bisa berjalan jika skenario melihat daftar pengajuan anggota sudah dijalankan. Proses lengkapnya dapat dilihat pada Lampiran C.

17) Melihat Detail Pengajuan Anggota

Activity Diagram melihat detail pengajuan anggota dapat dilakukan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menampilkan data diri pendaftar secara detail berdasarkan data pendaftaran. Proses lengkapnya dapat dilihat pada Lampiran C.

18) Menyetujui Pengajuan Anggota

Activity Diagram menyetujui pengajuan anggota dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam mengubah status pendaftar menjadi anggota dan mendapatkan akses API aplikasi. Proses lengkapnya dapat dilihat pada Lampiran C.

19) Menolak Pengajuan Anggota

Activity Diagram menolak pengajuan anggota dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan

aktor dalam menghapus data pendaftar dari basis data dan memberikan pemberitahuan terhadap pendaftar informasi penolakan pengajuan. Proses lengkapnya dapat dilihat pada Lampiran C.

20) Melihat Daftar Anggota

Activity Diagram melihat daftar anggota dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat menampilkan daftar anggota yang tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

21) Memfilter Daftar Anggota

Activity Diagram melihat daftar anggota dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat menampilkan daftar anggota yang tersimpan di dalam basis data berdasarkan filter yang dilakukan pengguna. Proses lengkapnya dapat dilihat pada Lampiran C.

22) Melihat Detai Anggota

Activity Diagram melihat detail anggota dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menampilkan detail anggota beserta log aktivitas API dari anggota tersebut. Proses lengkapnya dapat dilihat pada Lampiran C.

23) Menonaktifkan Keanggotaan

Activity Diagram menonaktifkan keanggotaan dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mengubah status keanggotaan anggota menjadi tidak aktif sekaligus menonaktifkan akses API dari anggota tersebut. Proses lengkapnya dapat dilihat pada Lampiran C.

24) Mengaktifkan Keanggotaan

Activity Diagram mengaktifkan keanggotaan dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mengubah status keanggotaan anggota menjadi aktor sekaligus mengaktifkan akses API dari anggota tersebut. Proses lengkapnya dapat dilihat pada Lampiran C.

25) Menghapus Keanggotaan

Activity Diagram menghapus keanggotaan dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menghapus data anggota dari basis data dan memberikan pemberitahuan terhadap pendaftar informasi penghapusan keanggotaan kepada anggota. Proses lengkapnya dapat dilihat pada Lampiran C.

26) Melihat Daftar Jenis Berita

Activity Diagram melihat daftar jenis berita dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar jenis berita yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

27) Menambahkan Jenis Berita Baru

Activity Diagram menambahkan jenis berita baru dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menambahkan jenis berita baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

28) Mengubah Jenis Berita

Activity Diagram mengubah jenis berita dapat dilakukan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mengubah data jenis berita yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

29) Menghapus Jenis Berita

Activity Diagram menghapus jenis berita dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menghapus data jenis berita yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

30) Melihat Daftar Pola Ekspresi Reguler

Activity Diagram melihat daftar pola ekspresi reguler dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar pola ekspresi reguler yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran AC

31) Menambahkan Pola Ekspresi Reguler

Activity Diagram menambahkan pola ekspresi reguler baru dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menambahkan pola ekspresi reguler baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

32) Mengubah Pola Ekspresi Reguler

Activity Diagram mengubah pola ekspresi reguler dapat dilakukan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mengubah data pola ekspresi reguler yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

33) Menghapus Pola Ekspresi Reguler

Activity Diagram menghapus pola ekspresi reguler dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menghapus data pola ekspresi reguler yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

34) Melihat Daftar Pola Khusus

Activity Diagram melihat daftar pola khusus dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar pola khusus yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran AC

35) Menambahkan Pola Khusus

Activity Diagram menambahkan pola khusus baru dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menambahkan pola khusus baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

36) Mengubah Pola Khusus

Activity Diagram mengubah pola khusus dapat dilakukan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mengubah data pola khusus yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

37) Menghapus Pola Khusus

Activity Diagram menghapus pola khusus dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menghapus data pola khusus yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

38) Melihat Daftar Filter HTML

Activity Diagram melihat daftar filter html dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor dalam menampilkan daftar filter html yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C

39) Menambahkan Pola Filter HTML

Activity Diagram menambahkan filter html baru dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menambahkan filter html baru dan menyimpannya ke dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

40) Mengubah Filter HTML

Activity Diagram mengubah filter html dapat dilakukan admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mengubah data filter html yang sebelumnya sudah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

41) Menghapus Filter HTML

Activity Diagram menghapus filter html dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas aplikasi pada saat berinteraksi dengan aktor untuk menghapus data filter html yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

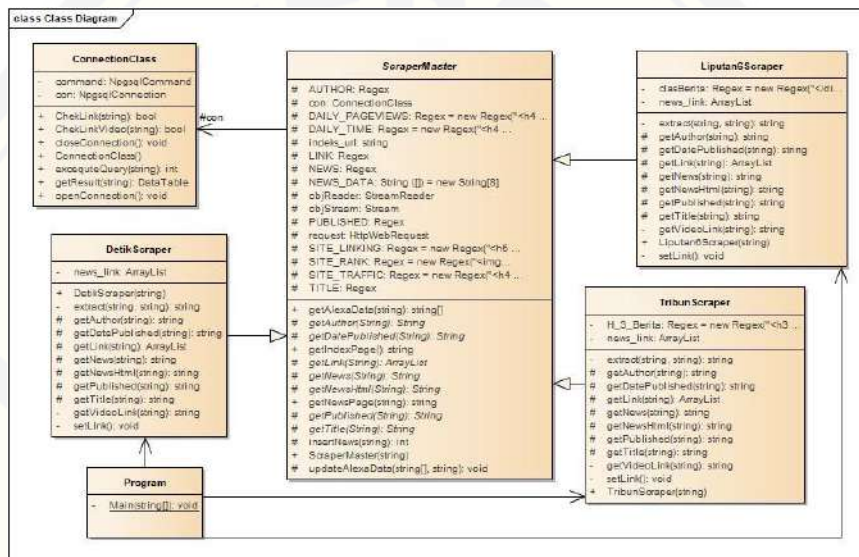
42) Mencari Judul Berita

Activity Diagram mencari judul berita dapat dilakukan oleh admin. *Activity Diagram* ini menjelaskan aktivitas-aktivitas aplikasi pada saat berinteraksi dengan aktor untuk mencari judul berita yang telah tersimpan di dalam basis data. Proses lengkapnya dapat dilihat pada Lampiran C.

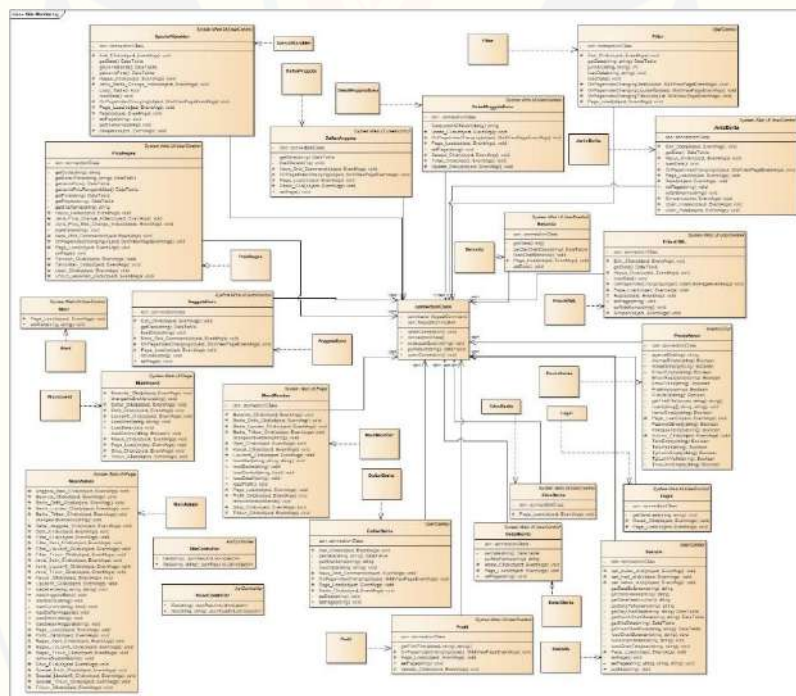
43) Class Diagram

Class Diagram menggambarkan struktur dan deskripsi *class* beserta

package yang memiliki hubungan satu sama lain dan digunakan untuk membangun aplikasi. Pembuatan *class diagram* sesuai dengan *class* apa saja yang dibutuhkan pada saat *website* aplikasi. *Class diagram* dari aplikasi yang dibangun dapat dilihat pada gambar 4.30 dan 4.31.



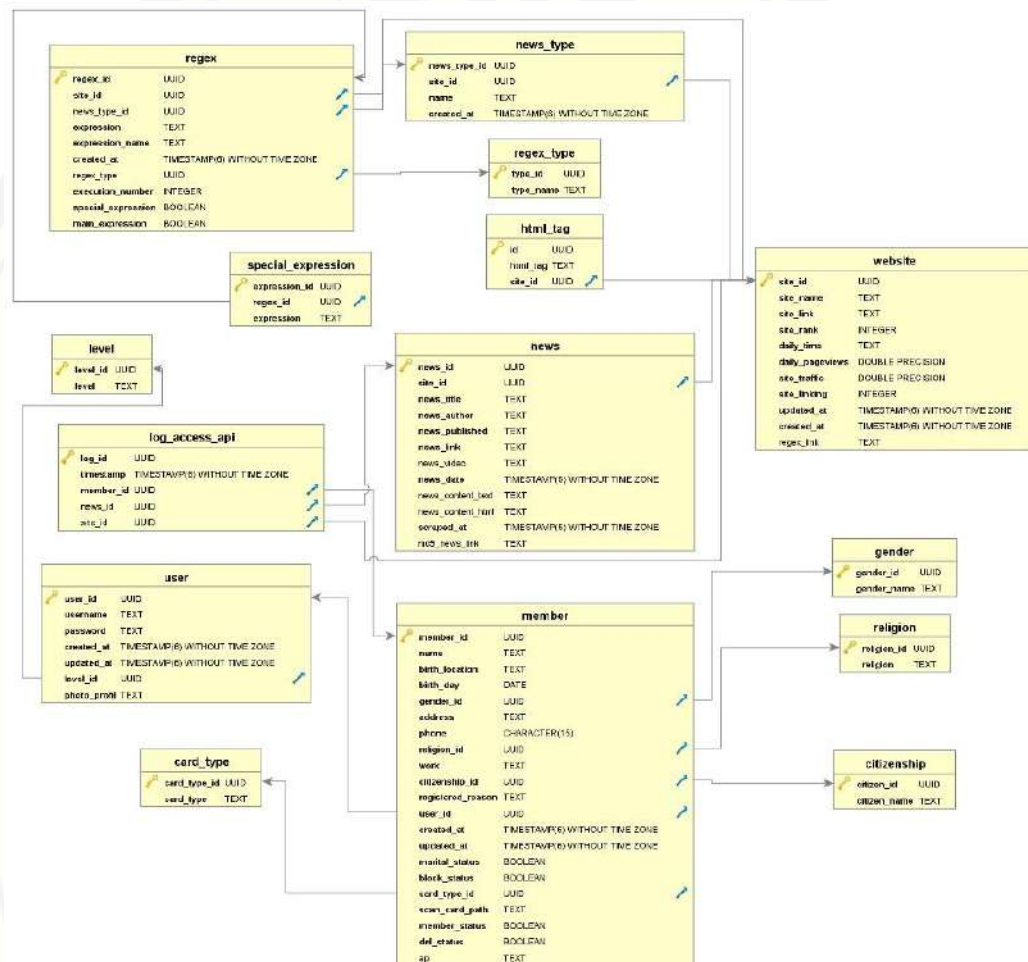
Gambar 4. 30 Class Diagram Scraper



Gambar 4. 31 Class Diagram Site Monitoring

4.4.6. Entity Relationship Diagram (ERD)

ERD merupakan gambaran komponen dan struktur *basis data* yang digunakan dalam pembangunan aplikasi. ERD pada aplikasi ekstraksi situs berita menggunakan metode ekspresi reguler dapat dilihat pada Gambar 4.31



Gambar 4. 32 Entiy Relationship Diagram

4.5. Implementasi Web Scraping

Setelah tahapan desain aplikasi selesai dilakukan, tahapan selanjutnya adalah implementasi aplikasi ke dalam bahasa *website*. Tahap implementasi dilakukan berdasarkan hasil desain aplikasi yang sudah dilakukan. Aplikasi yang dikembangkan dibagi menjadi dua sub aplikasi, yaitu *scarper* dan *monitoring*.

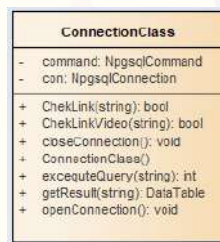
4.5.1. Scrapper

Scrapper adalah sub aplikasi yang dikembangkan untuk melakukan teknik

web scraping. Sub aplikasi ini dibangun menggunakan bahasa *website c#* dan dalam bentuk *Comand Line Interface* (CLI) seperti pada kebutuhan non-fungsional yang ada. Sub aplikasi ini akan dijalankan menggunakan *windows scheduler* setiap lima menit sekali untuk melakukan *web scraping* pada masing-masing situs. Pada sub aplikasi ini, diimplementasikan semua pola-pola *regex* yang ditemukan pada masing-masing situs. Berikut *class-class* yang dibangun pada saat pengembangan aplikasi.

1. *ConnectionClass*

Class ini dibangun untuk menangani koneksi dengan *basis data*. Pada *class* ini digunakan *driver library* dari *postgresql* berupa *Npgsql*. Peperincian *atribut* dan *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.33



Gambar 4. 33 *Class ConnectioClass*

Gambar 4.33 menggambarkan struktur *class ConnectionClass* yang dikembangkan. *Class* ini memiliki dua atribut, yaitu *command* dan *con*. Serta memiliki tujuh *method* di dalamnya.

2. *ScrapperMaster*

Class ini dibangun untuk menangani proses ekstraksi konten berita (melakukan *scraping*) secara umum. *Class* ini akan dijadikan *superclass* dari *class* khusus yang menangani proses *scraping* pada masing-masing situs. Peperincian atribut dan *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.34



Gambar 4. 34 *Class ScrapperMaster*

Gambar 4.34 menggambarkan struktur *class ScrapperMaster* yang dikembangkan. *Class* ini memiliki 16 *atribut*, serta memiliki 13 *method* di dalamnya.

3. *DetikScrapper*

Class ini dibangun khusus untuk menangani proses ekstraksi konten berita (melakukan *scraping*) pada situs detik.com. *Class* ini akan dijadikan *subclass* dari *class ScrapperMaster*. Perincian *atribut* dan *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.35.

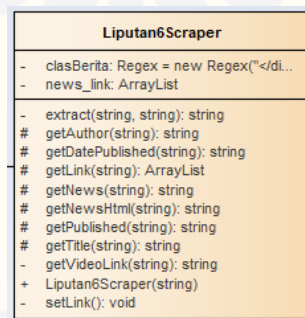


Gambar 4. 35 *Class DetikScrapper*

Gambar 4.35 menggambarkan struktur *class DetikScrapper* yang dikembangkan. *Class* ini memiliki satu *atribut*, yaitu *news_tautan* serta memiliki tiga *private method* dan tujuh *method* yang didapat dari *superclass*.

4. *Liputan6Scrapper*

Class ini dibangun khusus untuk menangani proses ekstraksi konten berita (melakukan *scraping*) pada situs liputan6.com. *Class* ini akan dijadikan *subclass* dari *class ScrapperMaster*. Perincian *atribut* dan *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.36.



Gambar 4. 36 *Class Liputan6Scrapper*

Gambar 4.36 menggambarkan struktur *class Liputan6Scraprer* yang dikembangkan. *Class* ini memiliki satu *atribut*, yaitu *news_tautan* dan *classBerita* serta memiliki tiga *private method* dan tujuh *method* yang didapat dari *superclass*.

5. *TribunScraprer*

Class ini dibangun khusus untuk menangani proses ekstraksi konten berita (melakukan *scraping*) pada situs *tribunnews.com*. *Class* ini akan dijadikan *subclass* dari *class ScrapperMaster*. Perincian *atribut* dan *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.37.

TribunScraprer	
-	H_3_Berita: Regex = new Regex("<h3 ...
-	news_link: ArrayList
-	extract(string, string): string
#	getAuthor(string): string
#	getDatePublished(string): string
#	getLink(string): ArrayList
#	getNews(string): string
#	getNewsHtml(string): string
#	getPublished(string): string
#	getTitle(string): string
-	getVideoLink(string): string
-	setLink(): void
+	TribunScraprer(string)

Gambar 4. 37 *Class TribunScraprer*

Gambar 4.37 menggambarkan struktur *class TribunScraprer* yang dikembangkan. *Class* ini memiliki satu *atribut*, yaitu *news_tautan* dan *H_3_Berita* serta memiliki tiga *private method* dan tujuh *method* yang didapat dari *superclass*.

6. *Program*

Class ini adalah *class* utama yang berfungsi untuk membentuk objek dari masing-masing *class* yang telah dibentuk sebelumnya. Perincian *class Program* dapat dilihat pada Gambar 4.38.

Program	
-	Main(string[]): void

Gambar 4. 38 *Class Program*

Gambar 4.38 menggambarkan struktur *class Program* yang dikembangkan. *Class* ini hanya memiliki satu *method* utama yang dijalankan ketika aplikasi dijalankan.

4.5.2. Monitoring

Monitoring adalah sub aplikasi yang dikembangkan sebagai *tools* untuk melakukan *monitoring* terhadap hasil ekstraksi yang dilakukan *scraper*. Sub aplikasi ini dibangun berbasis *web* menggunakan bahasa *website c#* dan *frame-work ASP.Net*. Pola-pola *regex* yang digunakan oleh *scraper* dapat diatur melalui sub aplikasi ini, sehingga memudahkan pengguna dalam mengatur pola apa saja yang akan digunakan oleh *scraper* agar dapat melakukan proses *scraping* dengan baik. Berikut *class-class* utama yang dibangun pada saat pengembangan aplikasi.

1. *connectionClass*

Class ini dibangun untuk menangani koneksi dengan *basis data*. Pada *class* ini digunakan *driver library* dari *postgresql* berupa *Npgsql*. Perincian *atribut* dan *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.39.

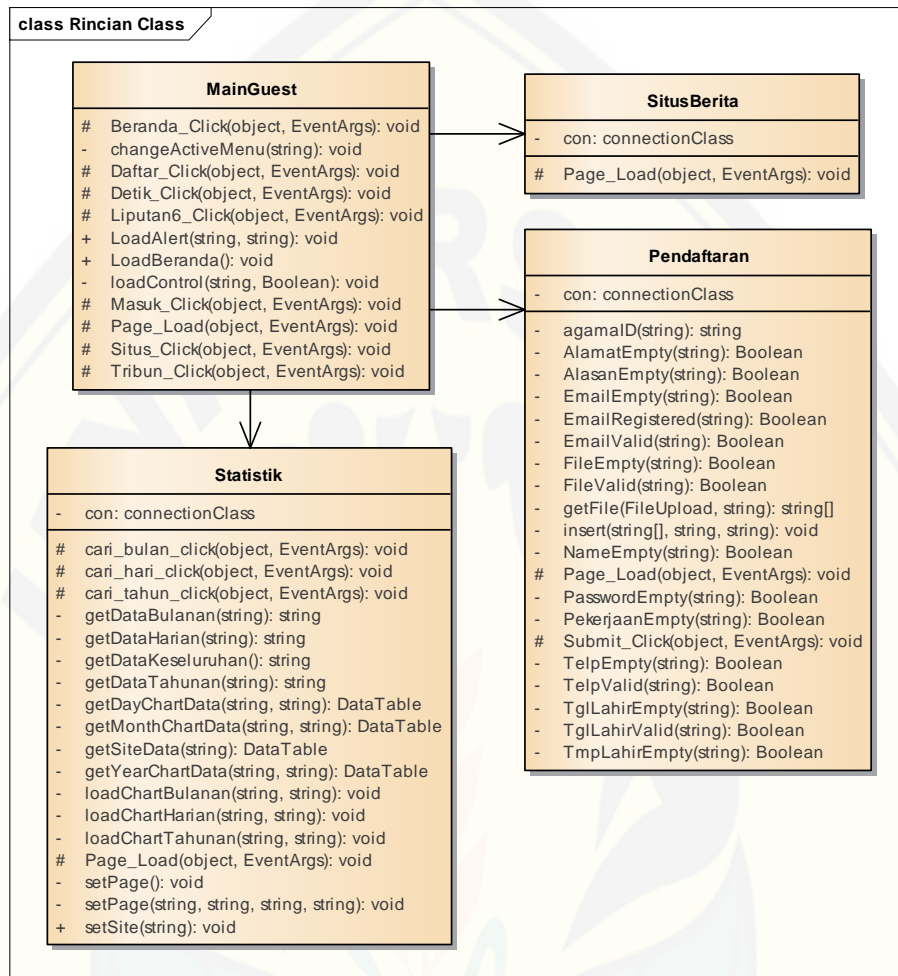
connectionClass	
-	command: NpgsqlCommand
-	con: NpgsqlConnection
-	closeConnection(): void
+	connectionClass()
+	excequiteQuery(string): void
+	getResult(string): DataTable
+	openConnection(): void

Gambar 4. 39 *Class connectionClass*

Gambar 4.39 menggambarkan struktur *class ConnectionClass* yang dikembangkan. *Class* ini memiliki dua atribut, yaitu *command* dan *con*. Serta memiliki lima *method* di dalamnya.

2. *MainGuest*

Class ini dibangun untuk menangani hak akses yang dimiliki oleh pengguna dengan level akses tamu. *Class* ini berhubungan langsung dengan *class-class* lain yang menangani fitur-fitur yang dapat diakses oleh tamu. Perincian *method* dan *class* apa saja yang berhubungan dengan *MainGuest* dapat dilihat pada Gambar 4.40.

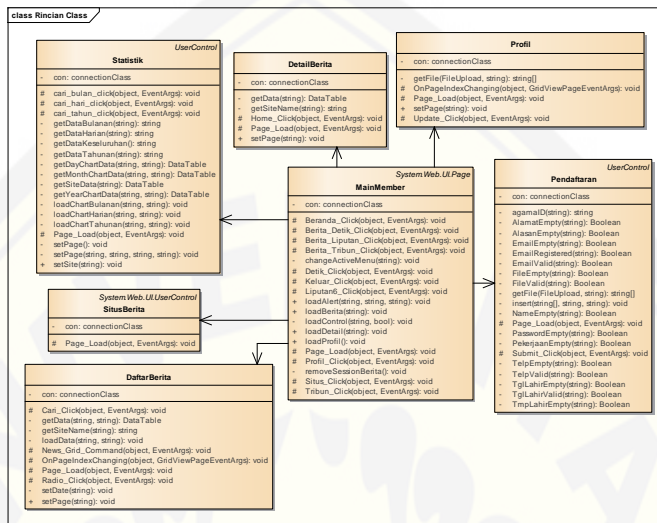


Gambar 4. 40 Class *MainGuest* dan relasinya

Gambar 4.40 menggambarkan struktur *class MainGuest* dan relasinya dengan *class* lain. Pada gambar tersebut dapat dilihat bahwa *class* ini memiliki delapan *protected method*, dua *private method*, dan dua *public method*. *Class* ini berhubungan secara langsung dengan *class* Pendaftaran, Statistik, dan SitusBerita.

3. *MainMember*

Class ini dibangun untuk menangani hak akses yang dimiliki oleh pengguna dengan level akses anggota. *Class* ini berhubungan langsung dengan *class-class* lain yang menangani fitur-fitur yang dapat diakses oleh anggota. Perincian *method* dan *class* apa saja yang berhubungan dengan *MainMember* dapat dilihat pada Gambar 4.41.

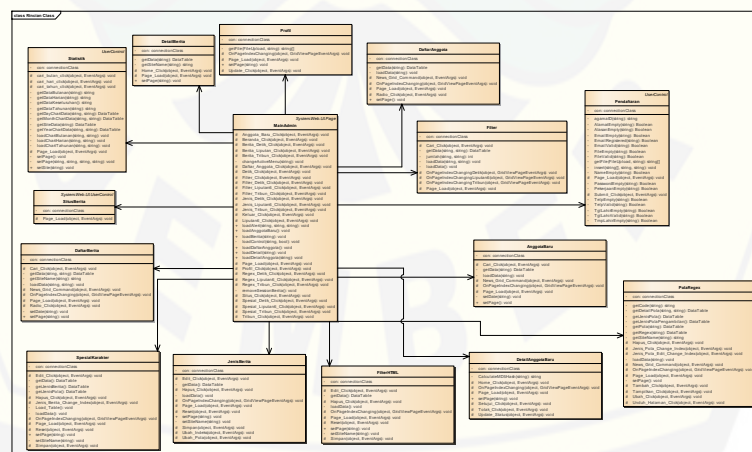


Gambar 4. 41 Class MainMember dan relasinya

Gambar 4.41 menggambarkan struktur *class MainMember* dan relasinya dengan *class* lain. Pada gambar tersebut dapat dilihat bahwa *class* ini memiliki 11 *protected method*, tiga *private method*, dan empat *public method*. *Class* ini berhubungan secara langsung dengan *class* Pendaftaran, Statistik, SitusBerita, DaftarBerita, DetailBerita dan Profil.

4. MainAdmin

Class ini dibangun untuk menangani hak akses yang dimiliki oleh pengguna dengan level akses admin. *Class* ini berhubungan langsung dengan *class-class* lain yang menangani fitur-fitur yang dapat diakses oleh admin. Perincian *method* dan *class* apa saja yang berhubungan dengan *MainAdmin* dapat dilihat pada Gambar 4.42.

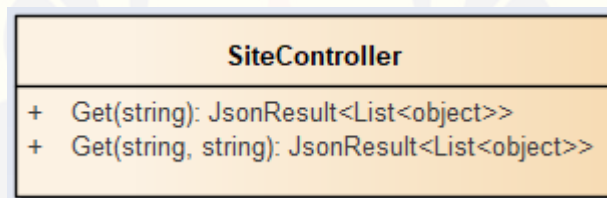


Gambar 4. 42 Class MainAdmin dan relasinya

Gambar 4.42 menggambarkan struktur *class MainAdmin* dan relasinya dengan *class* lain. Pada gambar tersebut dapat dilihat bahwa *class* ini memiliki 26 *protected method*, tiga *private method*, dan empat *public method*. *Class* ini berhubungan secara langsung dengan *class* Pendaftaran, Statistik, SitusBerita, DaftarBerita, DetailBerita, Profil, DaftarAnggota, AnggotaBaru, Filter, FilterHtml, DetailAnggotaBaru, PolaRegex, JenisBerita dan SpesialKarakter.

5. *SiteController*

Class ini dibangun untuk mengani layanan API yang disediakan oleh aplikasi. Data yang ditangani oleh *class* ini adalah data situs berita yang ada di dalam *basis data*. Perincian *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.43

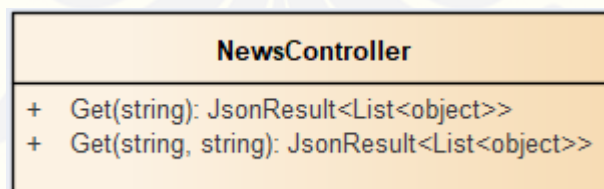


Gambar 4. 43 *Class SiteController*

Gambar 4.43 menggambarkan perincian *class SiteController* yang dibangun. *Class* ini memiliki dua *public method* dengan masing-masing mengembalikan data dalam bentuk *JsonResult*.

6. *NewsController*

Class ini dibangun untuk mengani layanan API yang disediakan oleh aplikasi. Data yang ditangani oleh *class* ini adalah data berita yang ada di dalam *basis data*. Perincian *method* yang dimiliki *class* ini dapat dilihat pada Gambar 4.44



Gambar 4. 44 *Class NewsController*

Gambar 4.44 menggambarkan perincian *class NewsController* yang dibangun. *Class* ini memiliki dua *public method* dengan masing-masing

mengembalikan data dalam bentuk *JsonResult*.

Setelah tahap implementasi selesai dilakukan, peneliti melakukan pengujian aplikasi menggunakan metode *black box*. Penguji yang melakukan pengujian aplikasi adalah salah satu mahasiswa yang sedang melakukan penelitian tentang analisis konten berita dengan menggunakan metode *text-mining*. Pengujian ini peneliti lakukan dua kali, pada pengujian pertama ditemukan dua kesalahan aplikasi yang tidak sesuai dengan skenario yang ada. Pada pengujian dua, tidak ditemukan kesalahan pada aplikasi, sehingga pengujian hanya dilakukan dua kali.

4.6. Uji Akurasi

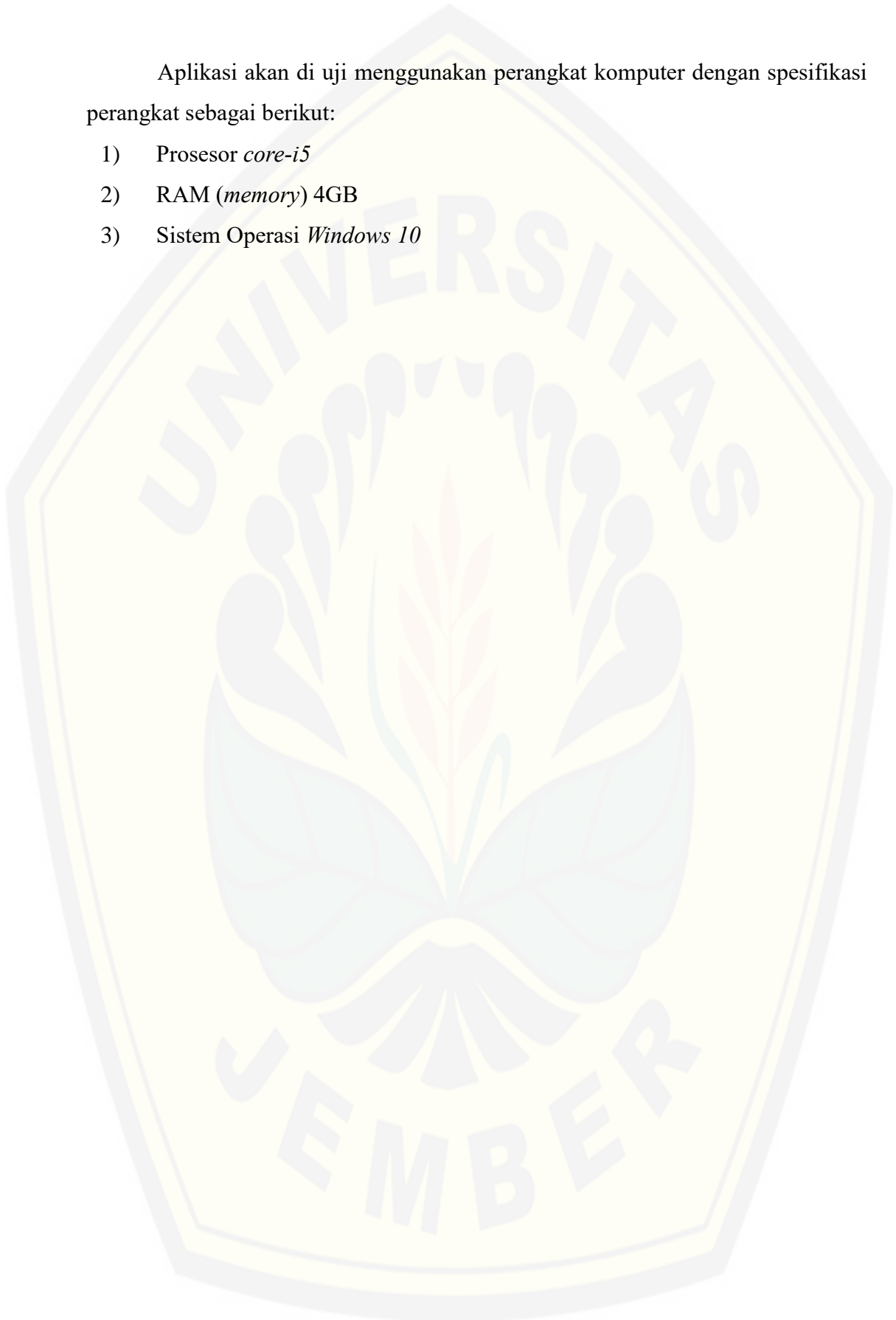
Untuk melakukan uji akurasi, peneliti menggunakan metode perbandingan *precision* dan *recall*, serta digabungkan dengan menggunakan perhitungan *f-measure*. Pada tahap ini peneliti melakukan uji akurasi berdasarkan data berita pada masing-masing situs pada tanggal 04 Juli 2018 pada jam 08.56 - 08.57 WIB. Pada tanggal tersebut, daftar berita yang ada pada masing-masing halaman indeks situs sejumlah 20 berita pada liputan6 dan tribunnuews, serta 21 berita pada detik.

Langkah pengujian dilakukan dengan melakukan proses *scraping* terhadap masing-masing situs. Setelah proses *scraping* selesai, peneliti melakukan pengecekan terhadap berapa jumlah data yang berhasil di ekstraksi, dan berapa jumlah data hasil ekstraksi yang sesuai dengan berita asli pada masing-masing situs. Tahap akhir uji akurasi adalah menghitung nilai *precision* dan *recall* dari masing-masing situs berdasarkan data yang sudah ada. Setelah nilai *precision* dan *recall* diketahui, maka dilakukan perhitungan *f-measure* dan akurasi dari masing-masing situs. Nilai akurasi yang didapatkan akan berbeda pada masing-masing situs, bergantung pada nilai *precision* dan *recall* masing-masing situs berita.

Relevant Document yang digunakan pada perhitungan *recall* dan *precision* berupa data asli dari ketiga situs yang diambil secara manual dari situs, sedangkan *Retrieved Document* adalah data berita yang berhasil diambil oleh *scraper*. Irisan dari kedua jenis dokumen ini berupa *Relevant Document Retrieved* adalah data berita yang berhasil di ekstraksi oleh *scraper* dan sesuai dengan data asli yang ada di masing-masing situs.

Aplikasi akan di uji menggunakan perangkat komputer dengan spesifikasi perangkat sebagai berikut:

- 1) Prosesor *core-i5*
- 2) RAM (*memory*) 4GB
- 3) Sistem Operasi *Windows 10*



BAB 6 PENUTUP

Bab ini berisi mengenai kesimpulan dan saran dari peneliti tentang penelitian yang telah dilakukan. Kesimpulan dan saran tersebut diharapkan dapat digunakan sebagai acuan pada penelitian selanjutnya.

6.1. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan peneliti, maka dapat disimpulkan beberapa kesimpulan berikut:

1. Berdasarkan hasil penelitian terhadap masing-masing situs berita, ditemukan beberapa pola *regex* yang dapat digunakan untuk melakukan proses ekstraksi konten dari halaman berita. Pola-pola yang ditemukan dapat dilihat pada Tabel 6.1

Tabel 6. 1 Pola *regex* yang ditemukan

Jenis Pola	Detik		Liputan6		Tribunnews	
	Video	Narasi	Video	Narasi	Video	Narasi
Tautan	<code><a data-category="Subkanal detikNews" data-action="Indeks" data-label="List Berita" href="(.*?)" class="list"></code>		<code></div>(.*?)<div class="article-snippet_wrapper-content"></code> <code></div></code>		<code><h3 class="blue">(.*?)</h3></code> <code><a href="(.*?)" title</code>	
Judul	<code><h1>(.*?)</h1></code>	<code><h1 class="\jd1\">(.*?)</h1></code>	<code><h1 class="\article-header_title\">(.*?)</h1></code>		<code><h1 class="\f32 fno crimson\" style="\line-height:40px;font-size:34px\">(.*?)</h1></code>	
Isi	<code><a(.*?)></code>	<code><!-- s:parallaxindetail-->(.*?)<!-- e:parallaxindetail--></code>	<code><div class="\seamless-ads_container\">(.*?)</div></div></code>		<code><div class='adspruce-bannerspot'(.*?)</div></code>	
	<code>(.*?)</code>	<code><table class="\tautansisip\">(.*?)</table></code> <code><table align="center" class="\pic_artikel_sisip_table\">(.*?)</table></code>	<code><div class="\seamless-ads\">(.*?)</div></code> <code><div class="\baca-juga\">(.*?)</div></code>		<code><span id='innity-in-post'(.*?)</code> <code><div id='div-Inside-MediumRectangle'(.*?)>(.*?)</div></code>	

Jenis Pola	Detik		Liputan6		Tribunnews	
	Video	Narasi	Video	Narasi	Video	Narasi
			<img(.*)>		<div class=\"txt-article mb20\" >(.*?)</div>	
		<div class=\"teks_detail detail_area\" id=\"detikdetailteks\" >(.*?)</div>	<div class=\"article-raw-content\" itemprop=\"description\" data-component-name=\"mobile:article-raw-content\">(.*?)</div>		Baca: <a(.*)	
		<iframe(.*)</iframe>	<a(.*)>		(.*?)	
		<a (.*)>	<blockquote(.*)100%;\">		<figure(.*)</figure>	
		<blockquote(.*)</blockquote>	<iframe class=\"vidio-embed\"(.*)</iframe>		<script(.*)</script>	
		<script(.*)</script>			<a(.*)>	
Penulis	(.*?)\t\t\t	<div class=\"author\">(.*?)</div>	(.*?)		<div class=\"f12 grey mb15\" >(.*?)</div>	
			>(.*?)			
Tanggal	20DETIK\t\t\t(.*?)\t\t</div>	<div class=\"date\">(.*?)</div>	(.*?)		<time class=\"grey f13 dip\">(.*?)</time>	
Tautan Video	<iframe src=\"(.*)\">	-	<iframe class=\"video-embed article-video-gallery-item-vidio-	-	<div class=\"video\" id=\"iframe-resize\"><iframe width=\"300\" height=\"169\" src=\"(.*)\">	-

Jenis Pola	Detik		Liputan6		Tribunnews	
	Video	Narasi	Video	Narasi	Video	Narasi
			embed\ " src="\ " (.*?)\ "			

2. Berdasarkan pola-pola *regex* yang ditemukan, terdapat perbedaan pola pada masing-masing situs berita. Perbedaan tersebut terletak pada *tag* apa yang digunakan, atribut apa saja yang dipakai, dan ada berapa pola *regex* yang dibutuhkan untuk mengekstrak satu jenis konten dari halaman situs berita.
3. Berdasarkan hasil uji akurasi yang dilakukan peneliti terhadap aplikasi yang dikembangkan, didapat hasil uji akurasi pada ketiga situs berita lebih besar dari 90%, yaitu Detik memiliki akurasi 100%, Liputan6 memiliki akurasi 100% dan Tribunnews memiliki akurasi 100%.

6.2. Saran

Berdasarkan hasil yang dicapai oleh peneliti dalam penelitian ini, berikut beberapa saran untuk penelitian selanjutnya.

1. Pola *regex* dalam penelitian ini masih dibentuk secara manual oleh peneliti, untuk penelitian selanjutnya akan lebih baik jika aplikasi yang dibentuk secara otomatis dapat membentuk pola *regex* tersendiri berdasarkan struktur html halaman yang diakses.
2. Agar data hasil ekstraksi dapat dimanfaatkan dengan baik, diperlukan penelitian lanjutan baik melalui teknik *text-mining* maupun *data-mining*.

DAFTAR PUSTAKA

- APJII. 2017. "Infografis Penetrasi & Perilaku Pengguna Internet Indonesia." Asosiasi Penyelenggara Jasa Internet Indonesia. <https://www.apjii.or.id/content/read/39/342/Hasil-Survei-Penetrasi-dan-Perilaku-Pengguna-Internet-Indonesia-2017>.
- ATR. 2017. "Top Sites in Indonesia." Alexa Traffic Rank. December 28, 2017. <https://www.alexa.com/topsites/countries/ID>.
- Cailliau, Robert. 2018. "A Little History of the World Wide Web." W3C. June 27, 2018. <https://www.w3.org/History.html>.
- Haddaway, Neal R. 2015. "The Use of Web-Scraping Software in Searching for Grey Literature." *Grey J* 11 (3): 186–90.
- Informatika, UPT Sumber Daya. 2017. "Belajar HTML – Konsep Dasar HTML." Belajar HTML – Konsep Dasar HTML. February 21, 2017. <http://si.usahid.ac.id/?p=1759>.
- Ling, Juen, I Putu Eka N. Kencana, and Tjokorda Bagus Oka. 2014. "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square." *E-Jurnal Matematika* 3 (3): 92. <https://doi.org/10.24843/MTK.2014.v03.i03.p070>.
- Mandala, Rila. 2006. "Evaluasi Efektifitas Metode Machine-Learning Pada Search-Engine." In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.
- Maria Rosario B, Yovi Pratama, and Fachruddin. 2014. "Penerapan Web Scraping Pada Website Company Profile." Sekolah Tinggi Ilmu Komputer Dinamika Bangsa.

<http://seminar.ilkom.unsri.ac.id/index.php/kntia/article/download/1173/57>.

Vivensius Mitra, and Dr Herry Sujaini. 2017. "Rancang Bangun Aplikasi Web Scraping Untuk Korpus Paralel Indonesia - Inggris Dengan Metode HTML DOM" 1 (1): 6.

Muliantara, Agus. 2009. "Penerapan Regular Expression Dalam Melindungi Alamat Email Dari Spam Robot Pada Konten Wordpress." *Jurnal Ilmu Komputer* 2 (1).

Prayitno, Agus. 2015. "Pemanfaatan Sistem Informasi Perpustakaan Digital Berbasis Website Untuk Para Penulis." *IJSE-Indonesian Journal on Software Engineering* 1 (1).

Utomo, Mardi Siswo. 2013. "Web Scraping Pada Situs Wikipedia Menggunakan Metode Ekspresi Regular." *Dinamik-Jurnal Teknologi Informasi* 18 (2).

Youssef Bassil. 2012. "A Simulation Model for the Waterfall Software Development Life Cycle" 2: 1-7.

LAMPIRAN

Lampiran A. Skenario

A.1. Melihat Beranda

Tabel A. 1 Skenario melihat beranda

No Usecase	USC2
Nama Usecase	Melihat beranda
Aktor	Tamu, Anggota, Admin
Deskripsi Singkat	Aktor melihat beranda aplikasi
Prakondisi	Aktor mengakses <i>tautan</i> aplikasi
Paskakondisi	Aplikasi menampilkan statistik <i>scraping</i> dari 3 situs dalam satu hari
Aktor	Aplikasi
Normal Flow	
1. Klik beranda	2. Menampilkan halaman beranda yang berisi <ul style="list-style-type: none"> - Grafik statistik <i>scraping</i> 3 situs dalam satu hari (sumbu y jumlah ekstraksi, sumbu x jam melakukan ekstraksi) - Informasi total hasil ekstraksi masing-masing situs dalam satu hari - Informasi total ekstraksi dari ketiga situs dalam satu hari

A.2. Mendaftar sebagai Anggota

Tabel A. 2 Skenario mendaftar sebagai anggota

No Usecase	USC1
Nama Usecase	Mendaftar sebagai anggota
Aktor	Tamu
Deskripsi Singkat	Aktor mengisi identitas pada formulir yang disediakan untuk proses pendaftaran sebagai anggota
Prakondisi	Aktor mengakses <i>tautan</i> aplikasi
Paskakondisi	Data identitas pendaftar tersimpan di dalam basis data dengan status menunggu persetujuan keanggotaan dari admin.
Aktor	Aplikasi
Normal Flow	
1. Klik menu daftar	2. Menampilkan halaman pendaftaran yang berisi formulir pendaftaran berupa : <ol style="list-style-type: none"> a. Informasi nomor telephone admin b. Pilihan jenis kartu identitas yang digunakan untuk mendaftar <ul style="list-style-type: none"> - KTP (<i>type: radio button, selected</i>) - SIM (<i>type: radio button</i>) c. Formulir Informasi akun <ul style="list-style-type: none"> - E-Mail (<i>type: e-mail, required</i>) - Password (<i>type: password, required</i>) - Foto profil (<i>type: file, jpg/jpeg, required</i>)

	<ul style="list-style-type: none"> d. Formulir Informasi Pendaftar <ul style="list-style-type: none"> - Nama (<i>type: teks, required</i>) - Tempat lahir (<i>type: teks, required</i>) - Tanggal lahir (<i>type: date, dd MM YYYY, required</i>) - Jenis kelamin (<i>type: radio button</i>) <ul style="list-style-type: none"> • Laki-laki (<i>selected</i>) • Perempuan - Alamat (<i>type: teksarea, required</i>) - Nomor telepon (<i>type: number, required</i>) - Agama (<i>type: select</i>) <ul style="list-style-type: none"> • Islam (<i>selected</i>) • Kristen protestan • Kristen katolik • Hindu • Budha • Konghucu - Status Perkawinan (<i>type : radio</i>) <ul style="list-style-type: none"> • Belum Kawin (<i>selected</i>) • Kawin - Pekerjaan (<i>type: teks, required</i>) - Kewarganegaraan (<i>type: radio</i>) <ul style="list-style-type: none"> • WNI (<i>selected</i>) • WNA - Alasan pengajuan (<i>type: teksarea, required</i>) - Bukti <i>scan</i> kartu identitas (<i>type: file, jpg/jpeg, required</i>) e. Tombol daftar (<i>type: submit</i>) f. Tombol batal (<i>type: reset</i>)
<ul style="list-style-type: none"> 3. Mengisi formulis pendaftaran 4. Klik daftar 	
	<ul style="list-style-type: none"> 5. Melakukan validasi dengan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menyimpan data pendaftar kedalam <i>basis data</i> dengan status menunggu persetujuan admin 8. Menampilkan pesan “Pendaftaran Sukses, Sedang Menunggu Persetujuan Admin” 9. <i>Redirect</i> ke halaman beranda.
<p style="text-align: center;">Alternatif flow : E-Mail tidak sesuai format / tidak terisi (validasi HTML)</p>	
<ul style="list-style-type: none"> 4. Klik daftar 	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
<p style="text-align: center;">Alternatif flow : E-Mail tidak terisi (validasi <i>back end</i>)</p>	
<ul style="list-style-type: none"> 4. Klik daftar 	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Alamat E-mail Anda !” 8. Tetap berada pada halaman pendaftaran

Alternatif flow : E-Mail tidak sesuai format (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Alamat E-mail Sesuai Format !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : E-Mail sudah terdaftar (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Alamat Email sudah terdaftar !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Password tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Password tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Password Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Foto profil tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Foto profil tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Foto untuk Profil Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Foto profil tidak sesuai format (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan File JPEG/JPG Sebagai Profil Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Nama tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Nama tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i>

	<ul style="list-style-type: none"> 7. Menampilkan pesan “Harap Masukkan Nama Lengkap Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Tempat lahir tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Tempat lahir tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Tempat Lahir Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Tanggal lahir tidak sesuai format / tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Tanggal lahir tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Tanggal Lahir Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Tanggal lahir tidak sesuai format (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Tanggal Lahir sesuai Format (dd MM YYYY) !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Alamat tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Alamat tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Malakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Alamat Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Nomor telepon tidak sesuai format / tidak terisi (validasi HTML)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Nomor telepon tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ul style="list-style-type: none"> 5. Melakukan validasi degan metode HTML <i>required</i>

	<ol style="list-style-type: none"> 6. Melakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Nomor Telepon Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Nomor telepon tidak sesuai format (validasi <i>back end</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Melakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Nomor Telepon berupa Angka !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Pekerjaan tidak terisi (validasi <i>HTML</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Pekerjaan tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Melakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Pekerjaan Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Alasan tidak terisi (validasi <i>HTML</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Alasan tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Melakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Alasan Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Bukti scan tidak terisi (validasi <i>HTML</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Menampilkan pesan error 7. Tetap berada pada halaman pendaftaran
Alternatif flow : Bukti scan tidak terisi (validasi <i>back end</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Melakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan Bukti Scan Kartu Identitas Anda !” 8. Tetap berada pada halaman pendaftaran
Alternatif flow : Bukti scan tidak sesuai format (validasi <i>back end</i>)	
4. Klik daftar	
	<ol style="list-style-type: none"> 5. Melakukan validasi degan metode <i>HTML required</i> 6. Melakukan validasi <i>back end</i> 7. Menampilkan pesan “Harap Masukkan File JPEG/JPG Pada Bukti Scan !” 8. Tetap berada pada halaman pendaftaran

Alternatif flow : Reset form	
4. Klik reset	
	5. Menghapus semua isian pada formulir pendaftaran.

A.3. Melihat Statistik Ekstraksi

Tabel A. 3 Skenario melihat detail statistik

No Usecase	USC2
Nama Usecase	Melihat detail statistik
Aktor	Tamu, Anggota, Admin
Deskripsi Singkat	Aktor melihat detail statistik ekstraksi dari salah satu situs berita
Prakondisi	Aktor mengakses <i>tautan</i> aplikasi
Paskakondisi	Aplikasi menampilkan detail statistik <i>scraping</i> dari salah satu situs berita.
Aktor	Aplikasi
Normal Flow	
1. Klik detail statistik	
	2. Menampilkan pilihan situs
3. Klik salah satu situs	
	<p>4. Menampilkan halaman detail statistik yang berisi :</p> <ol style="list-style-type: none"> Total hasil ekstraksi pada tanggal akses Total seluruh berita hasil ekstraksi dari situs Rata-rata jumlah berita yang dapat di ekstraksi per tahun. Tiga Grafik statistik : <ul style="list-style-type: none"> - Grafik statistik <i>scraping</i> situs harian (sumbu y jumlah ekstraksi, sumbu x jam melakukan ekstraksi) <ul style="list-style-type: none"> ➤ Rata-rata hasil ekstraksi perjam dalam satu hari ➤ Informasi prosentase harian dibandingkan dengan hasil ekstraksi satu bulan. $\frac{\text{Total Hasil Ekstraksi harian}}{\text{Total Hasil Ekstraksi Bulanan}} \times 100\%$ ➤ Informasi total hasil ekstraksi dalam satu bulan ➤ Form pencarian (<i>type: date, dd MM YYYY, required</i>) - Grafik statistik <i>scraping</i> situs bulanan (sumbu y jumlah ekstraksi, sumbu x tanggal melakukan ekstraksi) <ul style="list-style-type: none"> ➤ Rata-rata hasil ekstraksi perhari dalam satu bulan ➤ Informasi prosentase bulanan dibandingkan dengan hasil ekstraksi satu tahun. $\frac{\text{Total Hasil Ekstraksi Bulanan}}{\text{Total Hasil Ekstraksi Tahunan}} \times 100\%$ ➤ Informasi total hasil ekstraksi dalam satu tahun ➤ Form pencarian (<i>type: date, MM YYYY, required</i>) - Grafik statistik <i>scraping</i> situs tahunan (sumbu y jumlah ekstraksi, sumbu x bulan melakukan ekstraksi)

	<ul style="list-style-type: none"> ➤ Rata-rata hasil ekstraksi perbulan dalam satu tahun ➤ Informasi prosentase tahunan dibandingkan dengan hasil ekstraksi keseluruhan. $\frac{\text{Total Hasil Ekstraksi Tahunan}}{\text{Total Hasil Ekstraksi Keseluruhan}} \times 100\%$ ➤ Informasi total hasil ekstraksi keseluruhan ➤ Form pencarian (<i>type: date, YYYY, required</i>)
--	---

A.4. Mencari Detail Statistik

Tabel A. 4 Skenario mencari detail statistik

No Usecase	USC3
Nama Usecase	Mencari detail statistik
Aktor	Tamu, Anggota, Admin
Deskripsi Singkat	Aktor melakukan pencarian detail statistik pada waktu tertentu
Prakondisi	Aktor telah berada pada halaman detail statistik
Paskakondisi	Aplikasi menampilkan detail statistik <i>scraping</i> dari salah satu situs berita.
Aktor	Aplikasi
Normal flow	
1. Masukkan tanggal / bulan / tahun pada form pencarian	
2. Klik cari	
	<ol style="list-style-type: none"> 3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan hasil ekstraksi sesuai pencarian
Alternatif flow : Masukan tanggal tidak terisi (validasi <i>HTML</i>)	
2. Klik cari	
	<ol style="list-style-type: none"> 3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Menampilkan pesan eror 5. Tetap berada pada halaman detail statistik
Alternatif flow : Masukan tanggal tidak terisi (validasi <i>back end</i>)	
2. Klik cari	
	<ol style="list-style-type: none"> 3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan pesan "Masukkan Waktu pada Form Pencarian" 6. Tetap berada pada halaman detail statistik
Alternatif flow : Masukan tanggal tidak sesuai format (validasi <i>back end</i>)	
2. Klik cari	
	<ol style="list-style-type: none"> 3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan pesan "Masukkan Waktu Sesuai Format Pencarian" 6. Tetap berada pada halaman detail statistik

A.5. Melihat Daftar Situs Berita

Tabel A. 5 Skenario melihat daftar situs berita

No Usecase	USC4
Nama Usecase	Melihat daftar situs berita
Aktor	Tamu, Anggota, Admin
Deskripsi Singkat	Aktor melihat informasi tentang 3 situs berita.
Prakondisi	Aktor mengakses <i>tautan</i> aplikasi
Paskakondisi	Aplikasi menampilkan informasi tentang masing-masing situs berdasarkan data <i>real-time</i> dari www.alexacom.com
Aktor	Aplikasi
Normal Flow	
1. Klik menu situs berita	
	2. Menampilkan halaman situs berita yang berisi : <ol style="list-style-type: none"> a. Tanggal terakhir sinkronisasi dengan www.alexacom.com (<i>dd MM YYYY hh:mm:ss</i>) b. Informasi <i>Alexa Rank</i> masing-masing situs. c. Informasi <i>Daily Time</i> masing-masing situs (<i>MM:SS</i>) d. Informasi <i>Daily Page Views</i> masing-masing situs (<i>decimal</i>) e. Informasi <i>Traffic From Search</i> masing-masing situs (<i>prosentase</i>) f. Informasi <i>Site Tautaning</i> masing-masing situs

A.6. Masuk Aplikasi

Tabel A. 6 Skenario masuk aplikasi

No Usecase	USC5
Nama Usecase	Masuk
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor memasukkan email dan password untuk masuk kedalam aplikasi.
Prakondisi	Aktor mengakses <i>tautan</i> aplikasi
Paskakondisi	Aktor telah masuk kedalam aplikasi dan mendapatkan akses sesuai dengan hak akses masing-masing.
Aktor	Aplikasi
Normal Flow	
1. Klik masuk	
	2. Menampilkan halaman masuk aplikasi yang berisi : <ol style="list-style-type: none"> a. Masukkan e-mail (<i>type: email, required</i>) b. Masukkan password (<i>type: password, required</i>) c. Tombol masuk (<i>type: submit</i>)
3. Mengisi masukan email dan password.	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode <i>HTML required</i> 6. Melakukan validasi masukan di bagian <i>back end</i> 7. Melakukan validasi akun pengguna 8. Membuat <i>session</i> untuk pengguna

	9. Menampilkan halaman beranda serta hak akses lain sesuai dengan level pengguna
Alternatif flow : Email tidak sesuai format / tidak terisi (validasi HTML)	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode HTML <i>required</i> 6. Menampilkan pesan eror 7. Tetap berada pada halaman masuk aplikasi
Alternatif flow : Email tidak terisi (validasi <i>back end</i>)	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode HTML <i>required</i> 6. Melakukan validasi masukan di bagian <i>back end</i> 7. Menampilkan pesan “Silahkan Masukkan Alamat E-mail Anda !” 8. Tetap berada pada halaman masuk
Alternatif flow : Email tidak sesuai format (validasi <i>back end</i>)	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode HTML <i>required</i> 6. Melakukan validasi masukan di bagian <i>back end</i> 7. Menampilkan pesan “Silahkan Masukkan Alamat E-mail Sesuai Format !” 8. Tetap berada pada halaman masuk
Alternatif flow : Password tidak terisi (validasi HTML)	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode HTML <i>required</i> 6. Menampilkan pesan eror 7. Tetap berada pada halaman masuk aplikasi
Alternatif flow : Password tidak terisi (validasi <i>back end</i>)	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode HTML <i>required</i> 6. Melakukan validasi masukan di bagian <i>back end</i> 7. Menampilkan pesan “Silahkan Masukkan Password Anda !” 8. Tetap berada pada halaman masuk
Alternatif flow : Email / Password salah (validasi akun)	
4. Klik tombol masuk	
	5. Melakukan validasi masukan dengan menggunakan metode HTML <i>required</i> 6. Melakukan validasi masukan di bagian <i>back end</i> 7. Melakukan validasi akun pengguna 8. Menampilkan pesan “Maaf, Email atau Password yang Anda Masukkan Salah !” 9. Tetap berada pada halaman masuk

A.7. Melihat Daftar Berita

Tabel A. 7 Skenario melihat daftar berita

No Usecase	USC6
Nama Usecase	Melihat daftar berita
Aktor	Anggota, Admin

Deskripsi Singkat	Aktor melihat daftar berita yang berhasil di ekstraksi dari salah satu situs.
Prakondisi	Aktor telah masuk ke dalam aplikasi
Paskakondisi	Aplikasi menampilkan daftar berita yang berhasil di ekstraksi dari salah satu situs.
Aktor	Aplikasi
Normal Flow	
1. Klik menu daftar berita	
	2. Menampilkan pilihan situs berita
3. Klik salah satu situs	
	4. Menampilkan halaman daftar berita yang berisi : <ul style="list-style-type: none"> a. Informasi tanggal ekstraksi dari daftar berita yang ditampilkan. b. Form pencarian (<i>type: date, dd MM YYYY, required</i>) c. Tabel daftar berita tanpa <i>video</i> <ul style="list-style-type: none"> - No. - Judul (berisi judul berita) - Tautan (berisi tautan menuju halaman berita pada situs) - Aksi (berisi tombol detail yang berfungsi untuk melihat detail berita) d. Tabel daftar berita dengan <i>video</i> <ul style="list-style-type: none"> - Judul berita (berisi tautan ke halaman berita di situs) - Informasi berita (Penulis – Tanggal Publikasi) - Video (<i>type: iframe</i>) - Tombol detail e. Filter berita <ul style="list-style-type: none"> - Semua Berita, berfungsi menampilkan semua daftar berita (<i>type: radio, selected</i>) - Berita tanpa <i>video</i>, berfungsi hanya menampilkan berita tanpa <i>video</i> (<i>type: radio</i>) - Berita dengan <i>video</i>, berfungsi hanya menampilkan berita dengan <i>video</i> (<i>type: radio</i>)

A.8. Mencari Daftar Berita

Tabel A. 8 Skenario mencari daftar berita

No Usecase	USC7
Nama Usecase	Mencari daftar berita
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor melakukan pencarian terhadap daftar berita yang berhasil di ekstraksi dari salah satu situs berdasarkan waktu tertentu.
Prakondisi	Aktor telah berada pada halaman daftar berita
Paskakondisi	Aplikasi menampilkan daftar berita yang berhasil di ekstraksi dari salah satu situs berdasarkan waktu tertentu sesuai pencarian.
Aktor	Aplikasi
Normal Flow	

1. Masukkan tanggal / bulan / tahun pada form pencarian	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan berita hasil ekstraksi sesuai pencarian
Alternatif flow : Masukan tanggal tidak terisi (validasi <i>HTML</i>)	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Menampilkan pesan eror 5. Tetap berada pada halaman daftar berita
Alternatif flow : Masukan tanggal tidak terisi (validasi <i>back end</i>)	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan pesan "Masukkan Waktu pada Form Pencarian" 6. Tetap berada pada halaman daftar berita
Alternatif flow : Masukan tanggal tidak sesuai format (validasi <i>back end</i>)	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan pesan "Masukkan Waktu Sesuai Format Pencarian" 6. Tetap berada pada halaman daftar berita

A.9. Memfilter Daftar Berita

Tabel A. 9 Skenario memfilter daftar berita

No Usecase	USC8
Nama Usecase	Memfilter daftar berita
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor melakukan <i>filtering</i> terhadap daftar berita yang ditampilkan.
Prakondisi	Aktor telah berada pada halaman daftar berita
Paskakondisi	Aplikasi menampilkan daftar berita yang berhasil di ekstraksi dari salah satu situs berdasarkan filter yang dipilih.
Aktor	Aplikasi
Normal Flow	
1. Klik salah satu filter (semua berita, berita tanpa video, berita dengan video)	
	2. Menampilkan berita sesuai dengan filter yang dipilih pengguna

A.10. Melihat Detail Berita

Tabel A. 10 Skenario melihat detail berita

No Usecase	USC9
Nama Usecase	Melihat detail berita
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor melihat detail berita yang di pilih pada halaman daftar berita
Prakondisi	Aktor telah berada pada halaman daftar berita
Paskakondisi	Aplikasi menampilkan detail isi berita yang berhasil di ekstraksi dari salah satu situs berdasarkan berita yang dipilih pada daftar berita.
Aktor	Aplikasi
Normal Flow	
1. Klik detail pada salah satu daftar berita	
	2. Menampilkan halaman detail berita yang berisi : <ol style="list-style-type: none"> a. Judul berita b. Informasi publikasi berita <ul style="list-style-type: none"> - Penulis / editor - Tanggal publikasi c. Isi berita d. Informasi tanggal ekstraksi e. Tombol daftar berita yang berfungsi untuk kembali ke halaman daftar berita.

A.11. Melihat Profil

Tabel A. 11 Skenario melihat profil

No Usecase	USC10
Nama Usecase	Melihat profil
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor melihat profil akun yang dimiliki.
Prakondisi	Aktor telah masuk ke dalam aplikasi
Paskakondisi	Aplikasi menampilkan profil pengguna.
Aktor	Aplikasi
Normal Flow	
1. Klik profil	
	2. Menampilkan halaman profil yang berisi : <ol style="list-style-type: none"> a. Hak akses anggota berisi : <ul style="list-style-type: none"> - Foto profil - Identitas lengkap sesuai data pendaftaran - Alamat API masing-masing - Tanggal pendaftaran - Tombol ubah profil - Tabel <i>log</i> aktivitas API, yang berisi kolom <ul style="list-style-type: none"> ➤ No. ➤ Judul berita ➤ Tanggal akses

A.12. Mengubah Profil

Tabel A. 12 Skenario mengubah profil anggota

No Usecase	USC11
Nama Usecase	Mengubah profil
Aktor	Anggota
Deskripsi Singkat	Aktor melihat profil akun yang dimiliki.
Prakondisi	Aktor telah berada pada halaman profil
Paskakondisi	Aplikasi menampilkan profil pengguna.
Aktor	Aplikasi
Normal Flow	
1. Klik ubah	2. Menampilkan formulir ubah data dalam bentuk modal berisi: <ul style="list-style-type: none"> a. Formulir isian berupa: <ul style="list-style-type: none"> - Foto profil (<i>type: file</i>) - Password (<i>type: password</i>) b. Tombol simpan c. Tombol batal
3. Isi form isian	
4. Klik simpan	
	5. Simpan perubahan kedalam basis data
	6. Tampilkan pesan “data berhasil di simpan”
Alternatif flow: form isian kosong	
4. Klik simpan	
	5. Tidak melakukan perubahan data pada basis data.

Tabel A. 13 Skenario mengubah profil admin

No Usecase	USC11
Nama Usecase	Melihat profil
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor melihat profil akun yang dimiliki.
Prakondisi	Aktor telah masuk ke dalam aplikasi
Paskakondisi	Aplikasi menampilkan profil pengguna.
Aktor	Aplikasi
Normal Flow	
1. Klik profil	2. Menampilkan form ubah data dalam bentuk modal yang berisi: <ul style="list-style-type: none"> a. Formulir isian berupa: <ul style="list-style-type: none"> - Foto profil (<i>type: file</i>) - Username (<i>type: teks</i>) - Password (<i>type: password</i>) b. Tombol simpan c. Tombol batal
3. Isi form isian	
4. Klik simpan	
	5. Simpan perubahan kedalam basis data
	6. Tampilkan pesan “data berhasil disimpan”
Alternatif flow: form-isian kosong	
4. Klik simpan	
	5. Tidak melakukan perubahan data pada basis data

A.13. Keluar Aplikasi

Tabel A. 14 Skenario keluar aplikasi

No Usecase	USC12
Nama Usecase	Keluar
Aktor	Anggota, Admin
Deskripsi Singkat	Aktor keluar dari aplikasi.
Prakondisi	Aktor telah masuk ke dalam aplikasi
Paskakondisi	Aktor keluar dari aplikasi dan berada pada halaman beranda dengan akses tamu.
Aktor	Aplikasi
Normal Flow	
1. Klik keluar	
	<ol style="list-style-type: none"> 2. Menghapus semua <i>session</i> pengguna. 3. Menampilkan halaman beranda dengan hak akses tamu.

A.14. Melihat Daftar Pengajuan Anggota

Tabel A. 15 Skenario melihat daftar pengajuan anggota

No Usecase	USC13
Nama Usecase	Melihat daftar pengajuan anggota
Aktor	Admin
Deskripsi Singkat	Aktor melihat daftar pengajuan anggota yang telah mendaftar
Prakondisi	Aktor telah masuk ke dalam aplikasi
Paskakondisi	Aplikasi menampilkan daftar permintaan keanggotaan sesuai data pendaftar yang masuk
Aktor	Aplikasi
Normal Flow	
1. Klik menu anggota	
	<ol style="list-style-type: none"> 2. Menampilkan sub menu <ul style="list-style-type: none"> - Anggota baru - Daftar anggota
3. Klik anggota baru	
	<ol style="list-style-type: none"> 4. Menampilkan halaman daftar pengajuan keanggotaan yang berisi: <ol style="list-style-type: none"> a. Form pencarian berdasarkan waktu pendaftaran b. Tabel daftar pengajuan keanggotaan <ul style="list-style-type: none"> - No. - Nama - Alasan Pengajuan - Aksi (berisi tombol detail yang berfungsi untuk menampilkan detail pengajuan keanggotaan) c. Informasi jumlah pendaftar baru dalam satu hari sesuai data yang ditampilkan. d. Informasi jumlah anggota baru dalam satu bulan sesuai data yang ditampilkan e. Informasi jumlah anggota baru dalam satu tahun sesuai data yang ditampilkan

A.15. Mencari Daftar Pengajuan

Tabel A. 16 Skenario mencari daftar pengajuan

No Usecase	USC14
Nama Usecase	Mencari daftar pengajuan
Aktor	Admin
Deskripsi Singkat	Aktor melakukan pencarian terhadap daftar pengajuan yang masuk berdasarkan waktu tertentu.
Prakondisi	Aktor telah berada pada halaman daftar berita
Paskakondisi	Aplikasi menampilkan daftar pengajuan yang masuk berdasarkan waktu tertentu sesuai pencarian.
Aktor	Aplikasi
Normal Flow	
1. Masukkan tanggal / bulan / tahun pada form pencarian 2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan daftar pengajuan sesuai pencarian
Alternatif flow : Masukan tanggal tidak terisi (validasi HTML)	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Menampilkan pesan eror 5. Tetap berada pada halaman daftar pengajuan keanggotaan
Aternatif flow : Masukan tanggal tidak terisi (validasi <i>back end</i>)	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan pesan "Masukkan Waktu pada Form Pencarian" 6. Tetap berada pada halaman pengajuan keanggotaan
Aternatif flow : Masukan tanggal tidak sesuai format (validasi <i>back end</i>)	
2. Klik cari	
	3. Melakukan validasi masukan menggunakan metode <i>HTML required</i> 4. Melakukan validasi di bagian <i>back end</i> 5. Menampilkan pesan "Masukkan Waktu Sesuai Format Pencarian" 6. Tetap berada pada halaman pengajuan keanggotaan

A.16. Melihat Detai Pengajuan

Tabel A. 17 Skenario melihat detail pengajuan

No Usecase	USC15
Nama Usecase	Melihat detail pengajuan anggota
Aktor	Admin
Deskripsi Singkat	Aktor melihat detail salah satu pengajuan keanggotaan yang telah masuk
Prakondisi	Aktor telah berada pada halaman daftar pengajuan

Paskakondisi	Aplikasi menampilkan detail pengajuan sesuai dengan pengajuan yang dipilih.
Aktor	Aplikasi
Normal Flow	
1. Klik detail pada salah satu daftar pengajuan	
	2. Menampilkan halaman detail pengajuan yang berisi : <ol style="list-style-type: none"> a. Informasi pendaftar <ul style="list-style-type: none"> - Foto profil - Nama - Nomor identitas - Jenis kelamin - Tempat lahir - Tanggal lahir - Alamat - Status perkawinan - Pekerjaan - Kewarganegaraan - Nomor telepon - Email - Alasan pengajuan b. Jenis identitas yang digunakan c. Bukti scan kartu identitas d. Tanggal daftar e. Tombol Setujui f. Tombol Tolak

A.17. Menyetujui Pengajuan Anggota

Tabel A. 18 Skenario menyetujui keanggotaan

No Usecase	USC16
Nama Usecase	Menyetujui pengajuan keanggotaan
Aktor	Admin
Deskripsi Singkat	Aktor menyetujui pengajuan keanggotaan dan mengubah status pendaftar menjadi anggota
Prakondisi	Aktor telah berada pada halaman detail pengajuan
Paskakondisi	Aplikasi mengubah status pendaftar menjadi anggota dan menampilkan halaman daftar pengajuan anggota
Aktor	Aplikasi
Normal Flow	
1. Klik setuju	
	<ol style="list-style-type: none"> 2. Membuat API <i>read</i> data untuk anggota 3. Mengubah status pendaftar menjadi anggota 4. Menampilkan pesan "Pengajuan Berhasil Diterima" 5. Menampilkan halaman daftar pengajuan

A.18. Menolak Pengajuan Anggota

Tabel A. 19 Skenario menolak pengajuan anggota

No Usecase	USC17
Nama Usecase	Menolak pengajuan keanggotaan
Aktor	Admin
Deskripsi Singkat	Aktor menolak pengajuan keanggotaan dan menghapus data pendaftaran

Prakondisi	Aktor telah berada pada halaman detail pengajuan
Paskakondisi	Aplikasi menghapus data pendaftaran dan menampilkan halaman daftar pengajuan anggota
Aktor	Aplikasi
Normal Flow	
1. Klik Tolak	<ol style="list-style-type: none"> 2. Menghapus data pendaftar 3. Menampilkan pesan “Pengajuan Berhasil Ditolak” 4. Menampilkan halaman daftar pengajuan

A.19. Melihat Daftar Anggota

Tabel A. 20 Skenario melihat daftar anggota

No Usecase	USC18
Nama Usecase	Melihat daftar anggota
Aktor	Admin
Deskripsi Singkat	Aktor melihat daftar anggota yang mendapatkan akses API
Prakondisi	Aktor telah masuk ke dalam aplikasi
Paskakondisi	Aplikasi menampilkan daftar anggota yang mendapatkan akses API
Aktor	Aplikasi
Normal Flow	
1. Klik menu anggota	<ol style="list-style-type: none"> 2. Menampilkan sub menu <ul style="list-style-type: none"> - Anggota baru - Daftar anggota
3. Klik daftar anggota	<ol style="list-style-type: none"> 4. Menampilkan halaman daftar anggota yang berisi : <ol style="list-style-type: none"> a. Informasi Anggota Aktif <ul style="list-style-type: none"> - Jumlah anggota laki-laki - Jumlah anggota perempuan b. Informasi Anggota Nonaktif <ul style="list-style-type: none"> - Jumlah anggota laki-laki - Jumlah anggota perempuan c. Form pencarian (<i>data table</i>) d. Tabel daftar pengajuan keanggotaan <ul style="list-style-type: none"> - No. - Nama - Alasan Pengajuan - Aksi (berisi tombol detail yang berfungsi untuk menampilkan detail pengajuan keanggotaan) e. Filter (<i>radio button</i>) <ul style="list-style-type: none"> - Semua (<i>type: radio</i>) Berfungsi menampilkan semua anggota (aktif dan non aktif) - Aktif (<i>type: radio, selected</i>) Berfungsi menampilkan hanya anggota yang berstatus aktif - Nonaktif (<i>type: radio</i>) Berfungsi menampilkan hanya anggota yang berstatus nonaktif f. Informasi total jumlah anggota g. Informasi total anggota aktif

	h. Informasi total anggota nonaktif
--	-------------------------------------

A.20. Memfilter Daftar Anggota

Tabel A. 21 Skenario memfilter daftar anggota

No Usecase	USC19
Nama Usecase	Memfilter daftar anggota
Aktor	Admin
Deskripsi Singkat	Aktor melakukan <i>filtering</i> terhadap daftar anggota yang ditampilkan.
Prakondisi	Aktor telah berada pada halaman daftar anggota
Paskakondisi	Aplikasi menampilkan daftar anggota berdasarkan filter yang dipilih.
Aktor	Aplikasi
Normal Flow	
1. Klik salah satu filter (semu, aktif, nonaktif)	
	2. Menampilkan daftar anggota sesuai dengan filter yang dipilih pengguna

A.21. Melihat Detail Anggota

Tabel A. 22 Skenario melihat detail anggota

No Usecase	USC20
Nama Usecase	Melihat detail anggota
Aktor	Admin
Deskripsi Singkat	Aktor melihat detail salah satu anggota
Prakondisi	Aktor telah berada pada halaman daftar anggota
Paskakondisi	Aplikasi menampilkan detail anggota sesuai dengan anggota yang dipilih.
Aktor	Aplikasi
Normal Flow	
1. Klik detail pada salah satu daftar anggota	
	2. Menampilkan halaman detail anggota berisi : <ol style="list-style-type: none"> a. Informasi pendaftar <ul style="list-style-type: none"> - Foto profil - Nama - Nomor identitas - Jenis kelamin - Tempat lahir - Tanggal lahir - Alamat - Status perkawinan - Pekerjaan - Kewarganegaraan - Nomor telepon - Email - Alasan pengajuan b. Jenis identitas yang digunakan c. Bukti scan kartu identitas d. Tanggal menjadi anggota e. Tombol Bolkir (jika anggota aktif)

	f. Tombol Aktifkan (jika anggota nonaktif)
	g. Tombol hapus

A.22. Menonaktifkan Keanggotaan

Tabel A. 23 Skenario menonaktifkan keanggotaan

No Usecase	USC21
Nama Usecase	Menonaktifkan anggota
Aktor	Admin
Deskripsi Singkat	Aktor menonaktifkan keanggotaan salah satu anggota
Prakondisi	Aktor telah berada pada halaman detail anggota
Paskakondisi	Aplikasi mengubah status anggota menjadi tidak aktif dan meblokir akses API
Aktor	Aplikasi
Normal Flow	
1. Klik bolkir	
	<ol style="list-style-type: none"> 2. Mengubah status anggota menjadi tidak aktif 3. Menampilkan pesan “Anggota Dinonaktifkan” 4. Tetap berada pada halaman detail anggota dan mengubah tombol “blokir” menjadi “aktifkan”

A.23. Mengaktifkan Keanggotaan

Tabel A. 24 Skenario mengaktifkan keanggotaan

No Usecase	USC22
Nama Usecase	Mengaktifkan anggota
Aktor	Admin
Deskripsi Singkat	Aktor mengaktifkan keanggotaan salah satu anggota
Prakondisi	Aktor telah berada pada halaman detail anggota
Paskakondisi	Aplikasi mengubah status anggota menjadi aktif dan membuka akses API
Aktor	Aplikasi
Normal Flow	
1. Klik aktifkan	
	<ol style="list-style-type: none"> 2. Mengubah status anggota menjadi aktif 3. Menampilkan pesan “Anggota Diaktifkan” 4. Tetap berada pada halaman detail anggota dan mengubah tombol “aktifkan” menjadi “blokir”

A.24. Menghapus Keanggotaan

Tabel A. 25 Skenario menghapus keanggotaan

No Usecase	USC23
Nama Usecase	Menghapus anggota
Aktor	Admin
Deskripsi Singkat	Aktor menghapus data anggota dari <i>basis data</i>
Prakondisi	Aktor telah berada pada halaman detail anggota
Paskakondisi	Aplikasi menghapus data anggota dari <i>basis data</i>
Aktor	Aplikasi
Normal Flow	
1. Klik hapus	
	<ol style="list-style-type: none"> 2. Menampilkan pop up konfirmasi penghapusan anggota dengan pesan “Seluruh data tentang anggota

	ini akan dihapus ! Apakah anda yakin melanjutkan proses ?” dan pilihan “yakin” serta “tidak”
3. Klik yakin	
	<ol style="list-style-type: none"> 4. Menghapus seluruh data anggota dari dalam <i>basis data</i> 5. Menampilkan pesan “Data anggota berhasil di hapus” 6. Menampilkan halaman daftar anggota
Alternatif flow : batal hapus	
3. Klik tidak	
	4. Tetap berada pada halaman detail anggota

A.25. Melihat Daftar Jenis Berita

Tabel A. 26 Skenario melihat daftar jenis berita

No Usecase	USC24
Nama Usecase	Melihat daftar jenis berita
Aktor	Admin
Deskripsi Singkat	Aktor melihat daftar jenis berita dari salah satu situs.
Prakondisi	Aktor telah masuk ke dalam aplikasi dengan akses admin
Paskakondisi	Aplikasi menampilkan daftar jenis berita salah satu situs.
Aktor	Aplikasi
Normal Flow	
1. Klik menu jenis pola	
	2. Menampilkan pilihan situs berita
3. Klik salah satu situs	
	<ol style="list-style-type: none"> 4. Menampilkan halaman daftar jenis berita yang berisi: <ol style="list-style-type: none"> a. Form isian alamat indeks (langsung terisi dengan alamat indeks masing-masing situs sesuai basis data). b. Tombol ubah alamat indeks c. Form isian pola <i>regex</i> (langsung terisi dengan pola <i>regex</i> untuk ekstraksi tautan masing-masing situs sesuai dengan basis data) d. Tombol ubah pola tautan indeks. e. Tabel daftar jenis berita tanpa <ul style="list-style-type: none"> - No. - Nama Pola Berita (berisi nama pola) - Aksi (berisi tombol ubah dan tombol hapus) f. Form penambahan jenis berita baru <ul style="list-style-type: none"> - Nama jenis/pola berita (<i>type: teks</i>) - Tombol reset - Tombol simpan

A.26. Menambahkan Jenis Berita Baru

Tabel A. 27 Skenario menambah jenis berita baru

No Usecase	USC25
Nama Usecase	Menambahkan jenis berita baru
Aktor	Admin
Deskripsi Singkat	Aktor menambahkan jenis berita baru ke dalam basis data.
Prakondisi	Aktor telah berada pada halaman daftar jenis berita
Paskakondisi	Aplikasi menyimpan jenis berita baru ke dalam basis data.
Aktor	Aplikasi

Normal Flow	
1. Isi form penambahan jenis berita 2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menyimpan jenis berita baru ke dalam basis data. 6. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: nama pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: nama pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “nama pola tidak boleh kosong”

A.27. Mengubah Jenis Berita

Tabel A. 28 Skenario mengubah jenis berita

No Usecase	USC26
Nama Usecase	Mengubah jenis berita
Aktor	Admin
Deskripsi Singkat	Aktor mengubah data jenis berita yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar jenis berita
Paskakondisi	Aplikasi menyimpan perubahan data jenis berita dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu jenis berita 2. Klik ubah	
	3. Menampilkan nama jenis berita dari basis data pada form tambah data.
4. Ubah data 5. Klik Simpan	
	6. Melakukan validasi <i>front-end</i> 7. Melakukan validasi <i>back-end</i> 8. Menyimpan perubahan ke dalam basis data. 9. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: nama pola kosong (validasi <i>front-end</i>)	
5. Klik simpan	
	6. Melakukan validasi <i>front-end</i> 7. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: nama pola kosong (validasi <i>back-end</i>)	
5. Klik simpan	
	6. Melakukan validasi <i>front-end</i> 7. Melakukan validasi <i>back-end</i> 8. Menampilkan pesan “nama pola tidak boleh kosong”

A.28. Menghapus Jenis Berita

Tabel A. 29 Skenario menghapus jenis berita

No Usecase	USC27
Nama Usecase	Menghapus jenis berita
Aktor	Admin
Deskripsi Singkat	Aktor menghapus data jenis berita yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar jenis berita
Paskakondisi	Aplikasi menyimpan perubahan data jenis berita dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu jenis berita 2. Klik hapus	
	3. Menampilkan <i>alert</i> konfirmasi “Semua Pola dari Jenis Berita ini akan dihapus, apakah anda yakin ?”.
4. Klik ok	
	5. Menghapus data dari dalam basis data. 6. Menampilkan pesan “data berhasil dihapus”

A.29. Melihat Daftar Pola Ekspresi Regular

Tabel A. 30 Skenario melihat daftar pola regex

No Usecase	USC28
Nama Usecase	Melihat daftar pola <i>regex</i>
Aktor	Admin
Deskripsi Singkat	Aktor melihat daftar pola <i>regex</i> dari salah satu situs.
Prakondisi	Aktor telah masuk ke dalam aplikasi dengan akses admin
Paskakondisi	Aplikasi menampilkan daftar pola <i>regex</i> salah satu situs.
Aktor	Aplikasi
Normal Flow	
1. Klik menu pola <i>regex</i>	
	2. Menampilkan pilihan situs berita
3. Klik salah satu situs	
	4. Menampilkan halaman daftar pola <i>regex</i> yang berisi: <ul style="list-style-type: none"> a. Form masukan tautan halaman situs. b. Tombol unduh halaman. c. <i>Inspector</i> html d. Form penambahan pola <i>regex</i> baru <ul style="list-style-type: none"> - Jenis Berita (<i>type: dropdownlist</i>) - Jenis Konten (<i>type: dropdownlist</i>) - Urutan eksekusi (<i>type: dropdownlist</i>) - <i>Checklist</i> Membutuhkan Pola Khusus (<i>type: checkbox</i>) - <i>Checklist</i> Pola utama (<i>type: checkbox</i>) - Nama pola <i>regex</i> (<i>type: teks</i>) - Pola <i>regex</i> - Tombol simpan e. Pilihan jenis berita (<i>type: dropdownlist</i>) f. Tombol tampilkan g. Tabel daftar pola <i>regex</i> <ul style="list-style-type: none"> - No.

	<ul style="list-style-type: none"> - Nama Pola Berita (berisi nama pola) - Pola <i>regex</i> (berisi pola <i>regex</i>) - Pola pengambilan (berisi jenis konten) - Urutan eksekusi (berisi urutan eksekusi) - Aksi (berisi tombol ubah dan tombol hapus)
5. Pilih jenis berita 6. Klik tampilkan	
	<ul style="list-style-type: none"> 7. Tetap berada pada halaman daftar pola <i>regex</i> 8. Menampilkan semua pola <i>regex</i> dengan jenis berita yang dipilih pada tabel daftar pola <i>regex</i>.

A.30. Menambahkan Pola Ekspresi Regular

Tabel A. 31 Skenario menambah pola regex

No Usecase	USC29
Nama Usecase	Menambahkan pola <i>regex</i> baru
Aktor	Admin
Deskripsi Singkat	Aktor menambahkan pola <i>regex</i> baru ke dalam basis data.
Prakondisi	Aktor telah berada pada halaman daftar pola <i>regex</i>
Paskakondisi	Aplikasi menyimpan pola <i>regex</i> baru ke dalam basis data.
Aktor	Aplikasi
Normal Flow	
1. Isi form penambahan pola <i>regex</i> 2. Klik simpan	
	<ul style="list-style-type: none"> 3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menyimpan jenis berita baru ke dalam basis data. 6. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: nama pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	<ul style="list-style-type: none"> 3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: nama pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	<ul style="list-style-type: none"> 3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “nama pola tidak boleh kosong”
Alternatif flow: pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	<ul style="list-style-type: none"> 3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	<ul style="list-style-type: none"> 3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “pola tidak boleh kosong”

A.31. Mengubah Pola Ekspresi Regular

Tabel A. 32 Skenario mengubah pola regex

No Usecase	USC30
------------	-------

Nama Usecase	Mengubah pola <i>regex</i>
Aktor	Admin
Deskripsi Singkat	Aktor mengubah data pola <i>regex</i> yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar pola <i>regex</i>
Paskakondisi	Aplikasi menyimpan perubahan data pola <i>regex</i> dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu pola <i>regex</i> 2. Klik ubah	
	3. Menampilkan form ubah data pola yang sudah terisi data sesuai basis data yang berupa: <ul style="list-style-type: none"> - Jenis berita (<i>type: dropdownlist</i>) - Jenis konten (<i>type: dropdownlist</i>) - Urutan eksekusi (<i>type: dropdownlist</i>) - Nama pola (<i>type: teks</i>) - Pola <i>regex</i> (<i>type: teks</i>)
4. Ubah data 5. Klik Simpan	
	6. Melakukan validasi <i>front-end</i> 7. Melakukan validasi <i>back-end</i> 8. Menyimpan perubahan ke dalam basis data. 9. Menampilkan pesan “data berhasil disimpan” 10. Kembali menampilkan tabel daftar pola <i>regex</i>
Alternatif flow: nama pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: nama pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “nama pola tidak boleh kosong”
Alternatif flow: pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “pola tidak boleh kosong”

A.32. Menghapus Pola Ekspresi Reular

Tabel A. 33 Skenario menghapus pola *regex*

No Usecase	USC31
Nama Usecase	Menghapus pola <i>regex</i>
Aktor	Admin
Deskripsi Singkat	Aktor menghapus data pola <i>regex</i> yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar pola <i>regex</i>
Paskakondisi	Aplikasi menyimpan perubahan data pola <i>regex</i> dalam basis data

Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu pola <i>regex</i>	
2. Klik hapus	
	3. Menampilkan <i>alert</i> konfirmasi “Pola <i>regex</i> ini akan dihapus dari basis data secara permanen, apakah anda yakin ?”.
4. Klik ok	
	5. Menghapus data dari dalam basis data.
	6. Menampilkan pesan “data berhasil dihapus”

A.33. Melihat Daftar Pola Khusus

Tabel A. 34 Skenario melihat daftar Pola Khusus

No Usecase	USC32
Nama Usecase	Melihat daftar Pola Khusus
Aktor	Admin
Deskripsi Singkat	Aktor melihat daftar pola khusus dari salah satu situs.
Prakondisi	Aktor telah masuk ke dalam aplikasi dengan akses admin
Paskakondisi	Aplikasi menampilkan daftar pola khusus salah satu situs.
Aktor	Aplikasi
Normal Flow	
1. Klik menu Pola Khusus	
	2. Menampilkan pilihan situs berita
3. Klik salah satu situs	
	4. Menampilkan halaman daftar pola khusus yang berisi: <ol style="list-style-type: none"> a. Pilihan jenis berita (<i>type: dropdownlist</i>) b. Tabel daftar pola khusus <ul style="list-style-type: none"> - No. - Nama pola khusus (nama Pola Khusus) - Pola khusus untuk (pola <i>regex</i> yang membutuhkan karakter tersebut) - Aksi (berisi tombol ubah dan tombol hapus) c. Form penambahan jenis berita baru <ul style="list-style-type: none"> - Pilihan pola <i>regex</i> yang membutuhkan (<i>type: dropdownlist</i>) - Nama jenis/pola berita (<i>type: teks</i>) - Tombol reset - Tombol simpan

A.34. Menambahkan Pola Khusus

Tabel A. 35 Skenario menambahkan pola khusus

No Usecase	USC33
Nama Usecase	Menambahkan pola khusus baru
Aktor	Admin
Deskripsi Singkat	Aktor menambahkan pola khusus baru ke dalam basis data.
Prakondisi	Aktor telah berada pada halaman daftar pola khusus
Paskakondisi	Aplikasi menyimpan pola khusus baru ke dalam basis data.
Aktor	Aplikasi
Normal Flow	

1. Isi form penambahan pola khusus 2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menyimpan jenis berita baru ke dalam basis data. 6. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “nama pola tidak boleh kosong”

A.35. Mengubah Pola Khusus

Tabel A. 36 Skenario mengubah pola khusus

No Usecase	USC34
Nama Usecase	Mengubah pola khusus
Aktor	Admin
Deskripsi Singkat	Aktor mengubah data pola khusus yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar pola khusus
Paskakondisi	Aplikasi menyimpan perubahan data pola khusus dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu pola khusus 2. Klik ubah	
	3. Menampilkan pola khusus dari basis data pada form tambah data.
4. Ubah data 5. Klik Simpan	
	6. Melakukan validasi <i>front-end</i> 7. Melakukan validasi <i>back-end</i> 8. Menyimpan perubahan ke dalam basis data. 9. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: pola kosong (validasi <i>front-end</i>)	
8. Klik simpan	
	9. Melakukan validasi <i>front-end</i> 10. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: pola kosong (validasi <i>back-end</i>)	
9. Klik simpan	
	10. Melakukan validasi <i>front-end</i> 11. Melakukan validasi <i>back-end</i> 12. Menampilkan pesan “nama pola tidak boleh kosong”

A.36. Menghapus Pola Khusus

Tabel A. 37 Skenario menghapus pola khusus

No Usecase	USC35
Nama Usecase	Menghapus pola khusus
Aktor	Admin
Deskripsi Singkat	Aktor menghapus data pola khusus yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar pola khusus
Paskakondisi	Aplikasi menyimpan perubahan data pola khusus dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu pola khusus 2. Klik hapus	
	3. Menampilkan <i>alert</i> konfirmasi “Pola khusus akan dihapus dari dalam basis data, apakah anda yakin ?”.
4. Klik ok	
	5. Menghapus data dari dalam basis data. 6. Menampilkan pesan “data berhasil dihapus”

A.37. Melihat Daftar Filter HTML

Tabel A. 38 Skenario melihat daftar filter HTML

No Usecase	USC36
Nama Usecase	Melihat daftar filter html
Aktor	Admin
Deskripsi Singkat	Aktor melihat daftar filter html dari salah satu situs.
Prakondisi	Aktor telah masuk ke dalam aplikasi dengan akses admin
Paskakondisi	Aplikasi menampilkan daftar filter html salah satu situs.
Aktor	Aplikasi
Normal Flow	
1. Klik menu filter html	
	2. Menampilkan pilihan situs berita
3. Klik salah satu situs	
	4. Menampilkan halaman daftar filter html yang berisi: <ol style="list-style-type: none"> a. Tabel daftar filter html tanpa <ul style="list-style-type: none"> - No. - Ekspresi (ekspresi) - Aksi (berisi tombol ubah dan tombol hapus) b. Form penambahan jenis berita baru <ul style="list-style-type: none"> - Expresi (<i>type: teks</i>) - Tombol reset - Tombol simpan

A.38. Menambahkan Filter HTML Baru

Tabel A. 39 Skenario menambahkan filter html baru

No Usecase	USC37
Nama Usecase	Menambahkan filter html baru
Aktor	Admin
Deskripsi Singkat	Aktor menambahkan filter html baru ke dalam basis data.
Prakondisi	Aktor telah berada pada halaman daftar filter html

Paskakondisi	Aplikasi menyimpan filter html baru ke dalam basis data.
Aktor	Aplikasi
Normal Flow	
1. Isi form penambahan filter html 2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menyimpan jenis berita baru ke dalam basis data. 6. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: ekspresi pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: ekspresi pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “ekspresi tidak boleh kosong”

A.39. Mengubah Filter HTML

Tabel A. 40 Skenario mengubah filter html

No Usecase	USC38
Nama Usecase	Mengubah filter html
Aktor	Admin
Deskripsi Singkat	Aktor mengubah data filter html yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar filter html
Paskakondisi	Aplikasi menyimpan perubahan data filter html dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu filter html 2. Klik ubah	
	3. Menampilkan filter html dari basis data pada form tambah data.
4. Ubah data 5. Klik Simpan	
	6. Melakukan validasi <i>front-end</i> 7. Melakukan validasi <i>back-end</i> 8. Menyimpan perubahan ke dalam basis data. 9. Menampilkan pesan “data berhasil disimpan”
Alternatif flow: ekspresi pola kosong (validasi <i>front-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Menampilkan pesan error (<i>html required</i>)
Alternatif flow: ekspresi pola kosong (validasi <i>back-end</i>)	
2. Klik simpan	
	3. Melakukan validasi <i>front-end</i> 4. Melakukan validasi <i>back-end</i> 5. Menampilkan pesan “ekspresi pola tidak boleh kosong”

A.40. Menghapus Filter HTML

Tabel A. 41 Skenario menghapus filter html

No Usecase	USC39
Nama Usecase	Menghapus filter html
Aktor	Admin
Deskripsi Singkat	Aktor menghapus data filter html yang ada pada basis data.
Prakondisi	Aktor telah berada pada halaman daftar filter html
Paskakondisi	Aplikasi menyimpan perubahan data filter html dalam basis data
Aktor	Aplikasi
Normal Flow	
1. Pilih salah satu filter html 2. Klik hapus	
	3. Menampilkan <i>alert</i> konfirmasi “Filter html ini akan dihapus, apakah anda yakin ?”.
4. Klik ok	
	5. Menghapus data dari dalam basis data. 6. Menampilkan pesan “data berhasil dihapus”

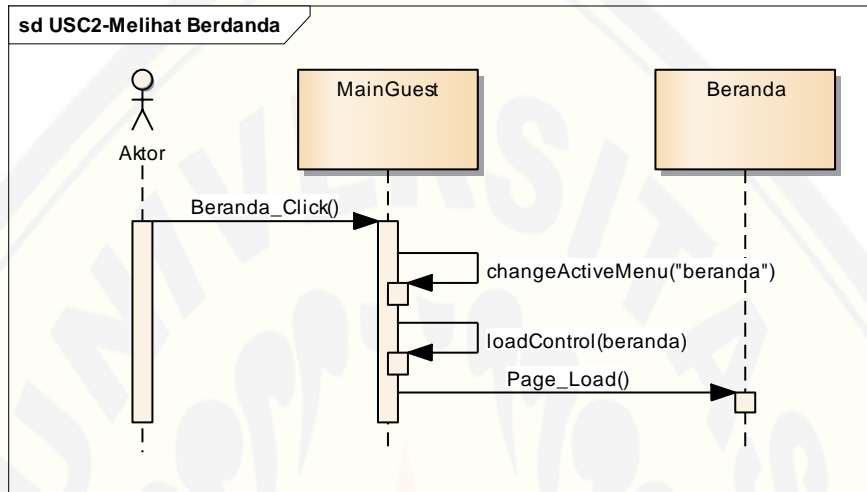
A.41. Mencari judul berita

Tabel A. 42 Skenario mencari judul berita

No Usecase	USC40
Nama Usecase	Mencari judul berita
Aktor	Admin
Deskripsi Singkat	Aktor mencari judul berita dari data berita yang sudah berhasil di ekstraksi.
Prakondisi	Aktor telah masuk ke dalam aplikasi dengan akses admin
Paskakondisi	Aplikasi menampilkan berita berdasarkan kata kunci yang dimasukkan.
Aktor	Aplikasi
Normal Flow	
1. Klik menu berita	
	2. Menampilkan pilihan filter berita
3. Klik filter berita	
	4. Menampilkan halaman filter berita yang berisi: a. Form isian judul berita (<i>type: teks</i>) b. Tombol cari judul berita
5. Masukkan kata kunci 6. Klik cari judul berita	
	7. Menampilkan hasil pencarian berdasarkan kata kunci dalam bentuk tabel masing-masing situs. - No. - Judul Berita 8. Menampilkan jumlah dan prosentase jumlah berita yang ditemukan pada masing-masing situs.

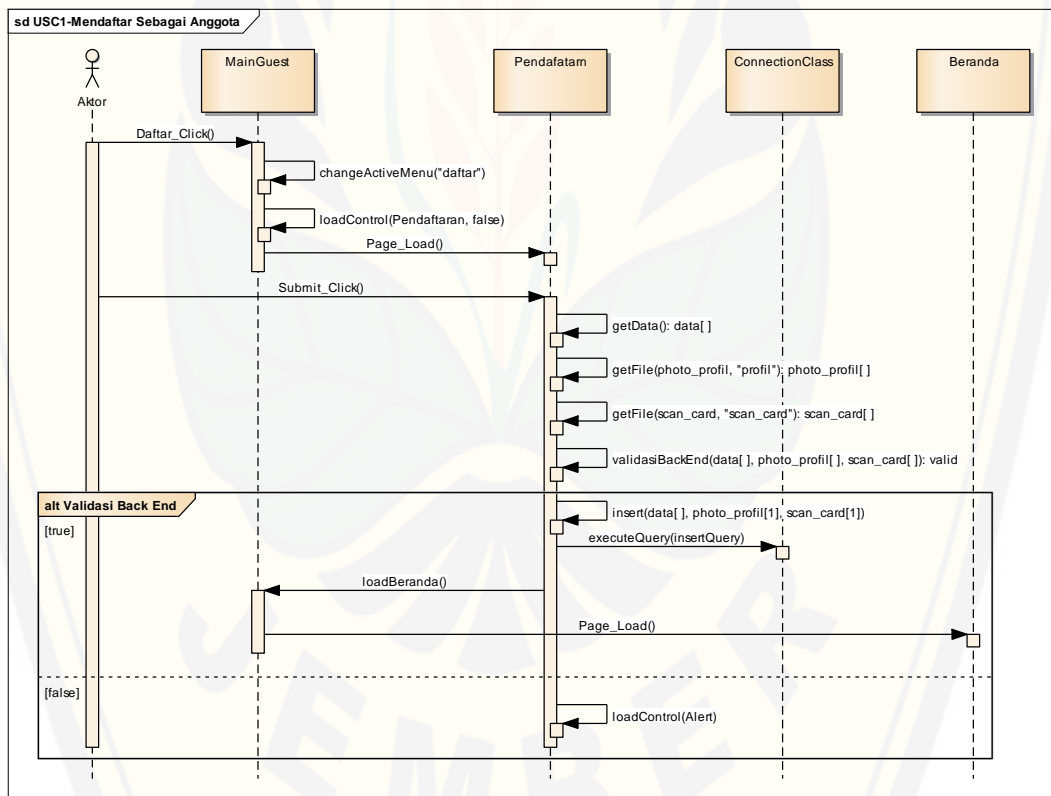
Lampiran B. Sequence Diagram

B.1. Melihat Beranda



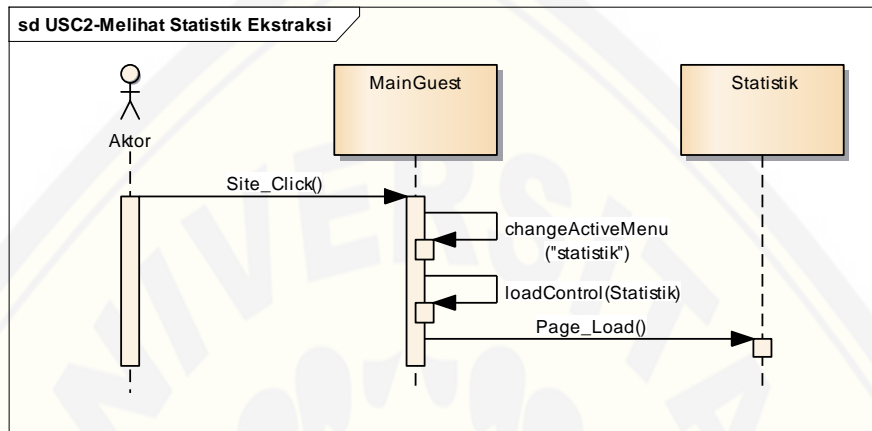
Gambar B. 1 Sequence diagram melihat beranda

B.2. Mendaftar sebagai Anggota



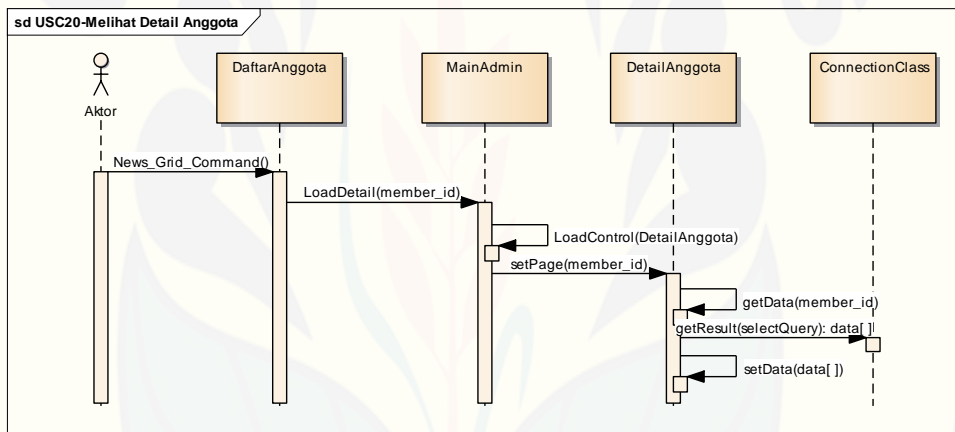
Gambar B. 2 Sequence diagram mendaftar sebagai anggota

B.3. Melihat Statistik Ekstraksi



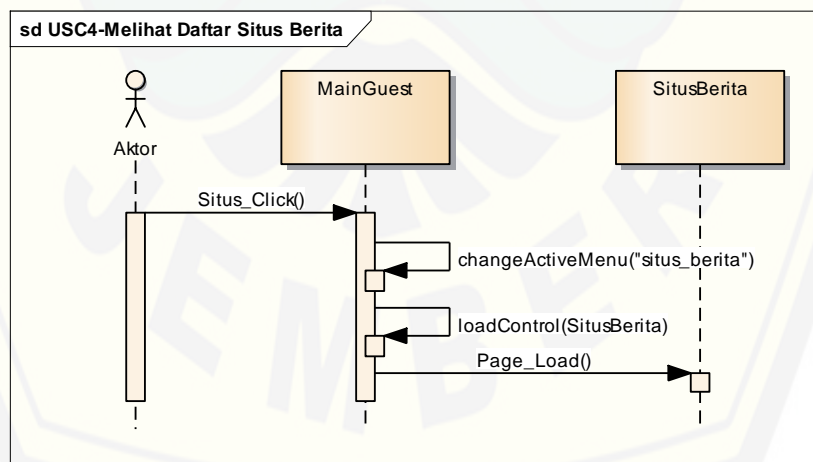
Gambar B. 3 Sequence diagram melihat statistik ekstraksi

B.4. Mencari Detail Statistik



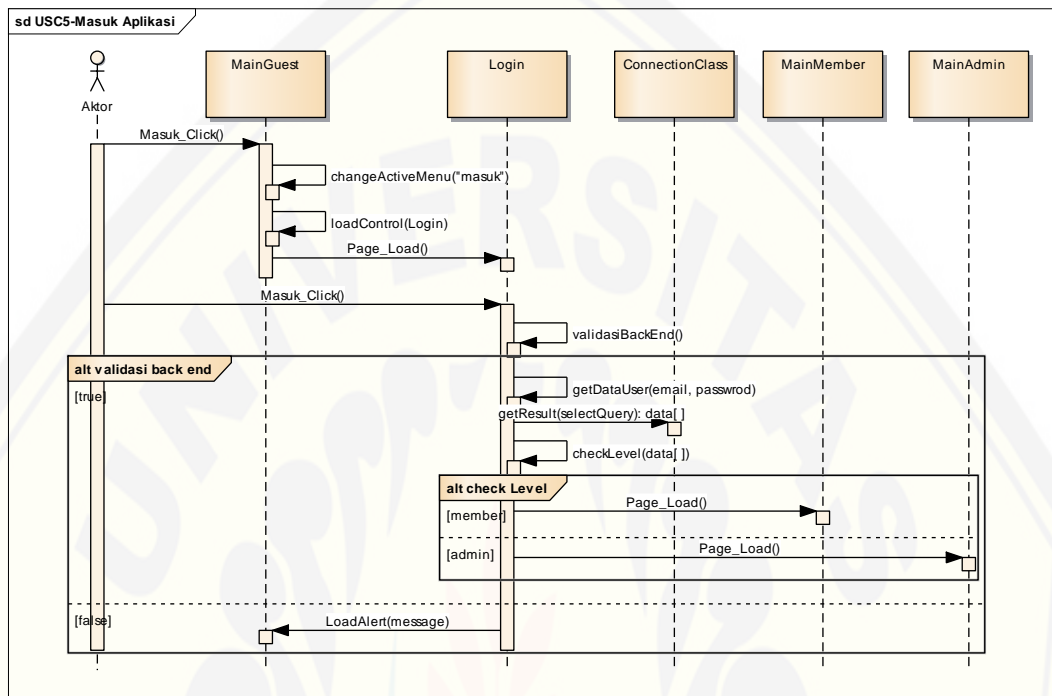
Gambar B. 4 Sequence diagram detail statistik

B.5. Melihat Daftar Situs Berita



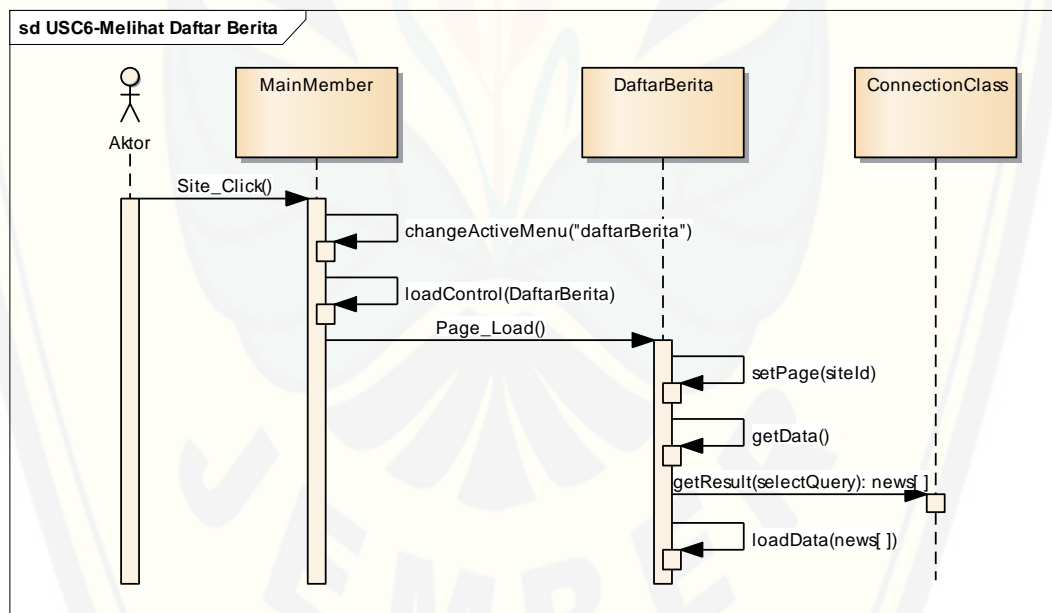
Gambar B. 5 Sequence diagram melihat daftar situs berita

B.6. Masuk Aplikasi



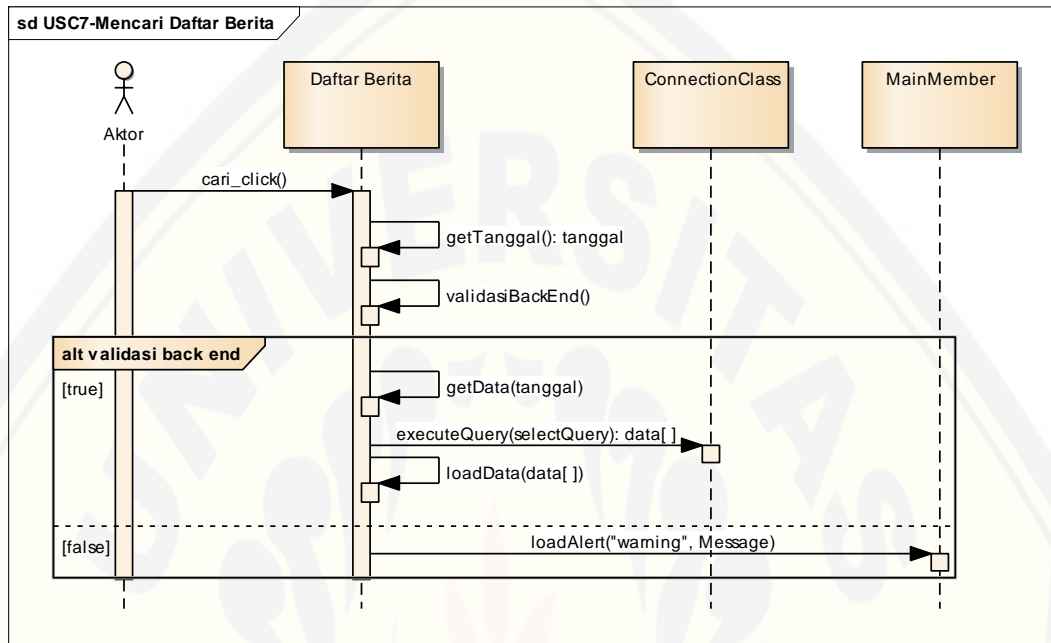
Gambar B. 6 Sequence diagram masuk aplikasi

B.7. Melihat Daftar Berita



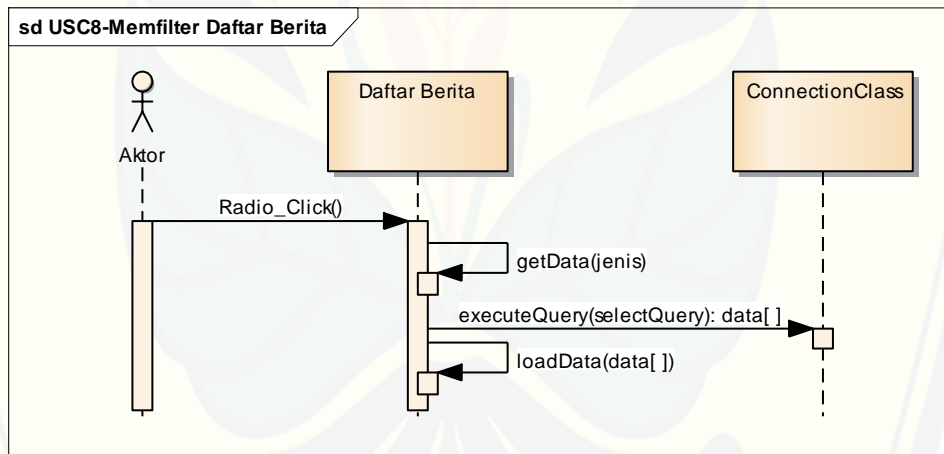
Gambar B. 7 Sequence diagram melihat daftar berita

B.8. Mencari Daftar Berita



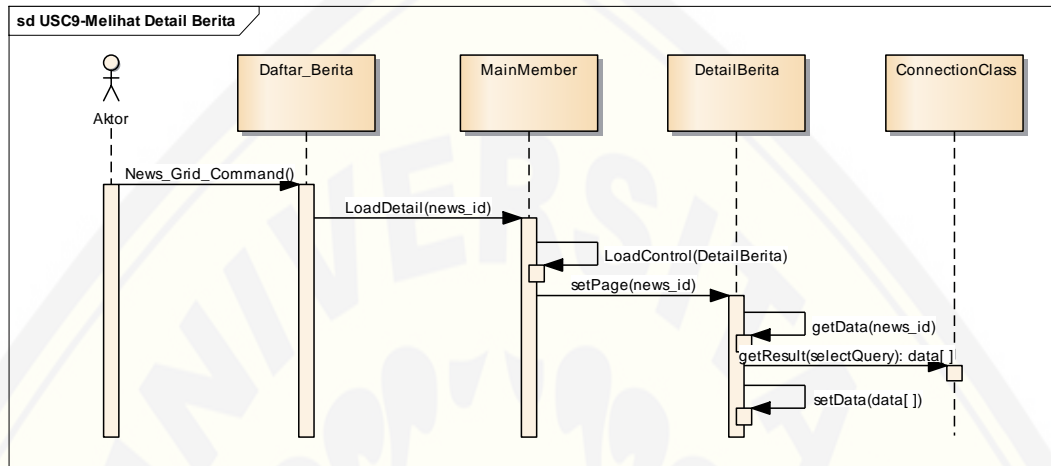
Gambar B. 8 Sequence diagram mencari daftar berita

B.9. Memfilter Daftar Berita



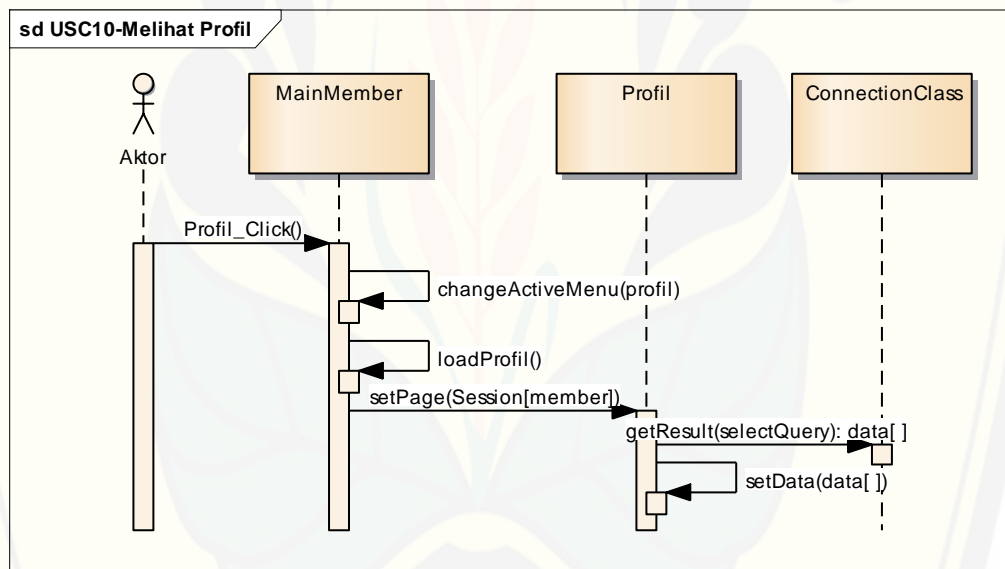
Gambar B. 9 Sequence diagram memfilter daftar berita

B.10. Melihat Detail Berita



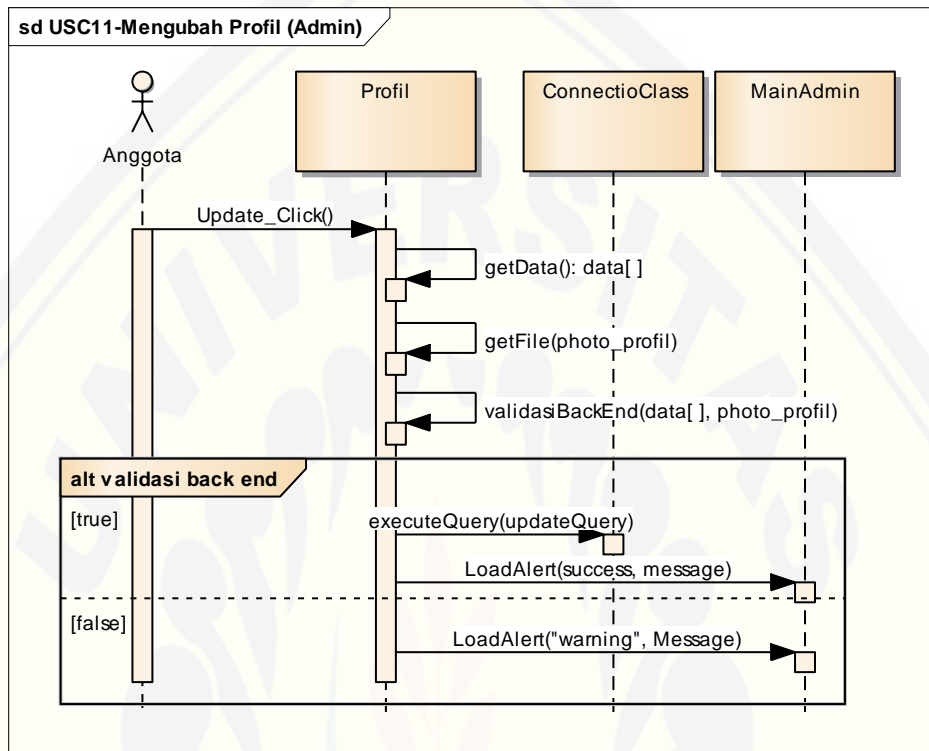
Gambar B. 10 Sequence diagram melihat detail berita

B.11. Melihat Profil

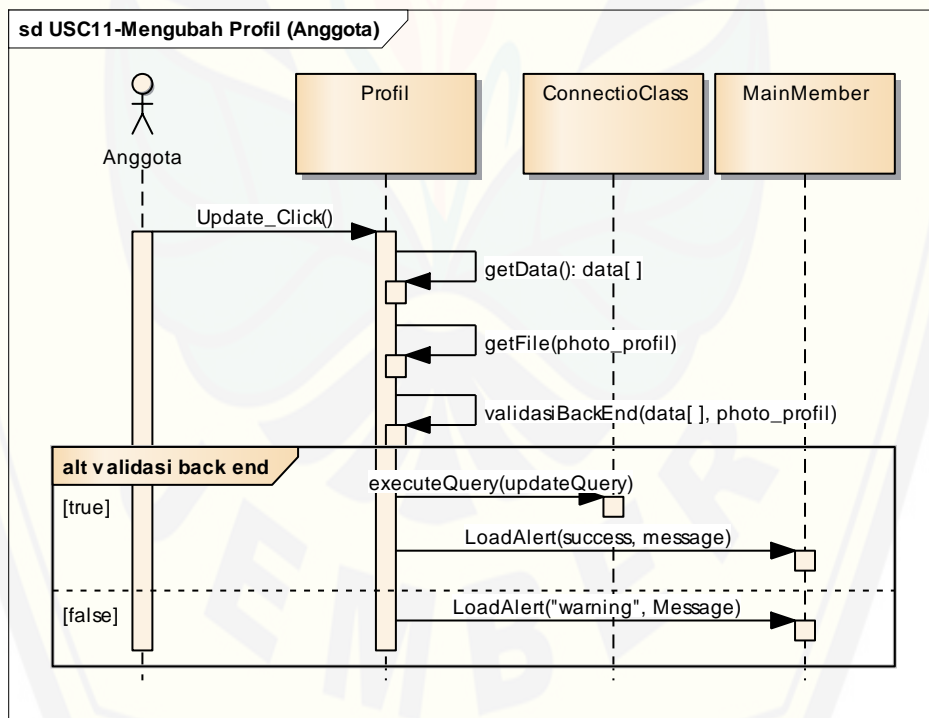


Gambar B. 11 Sequence diagram melihat profil

B.12. Mengubah Profil

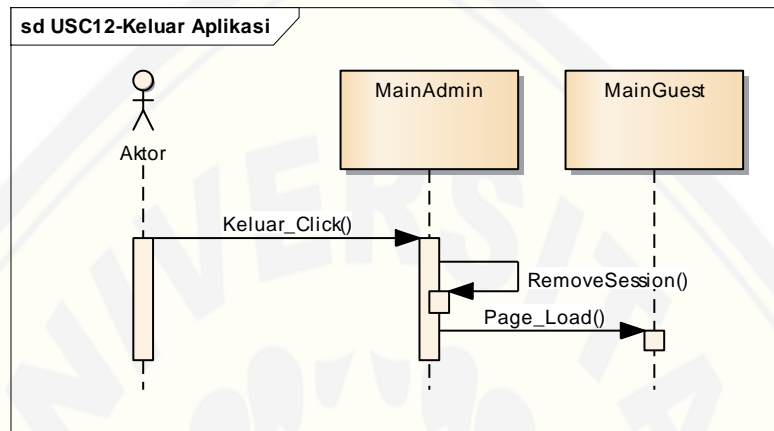


Gambar B. 12 Sequence diagram mengubah foto profil (admin)



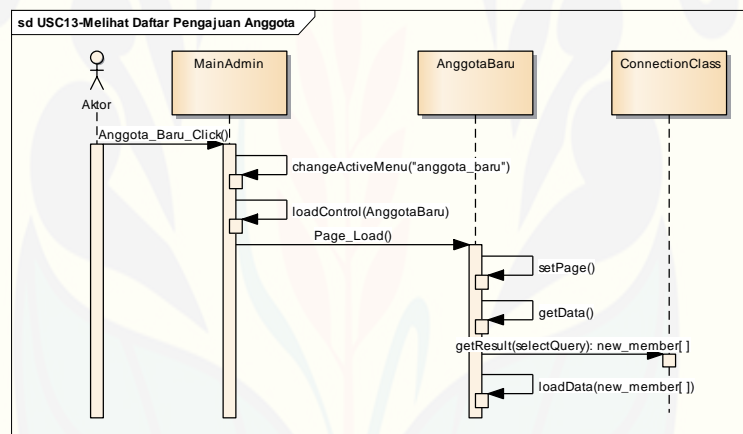
Gambar B. 13 Sequence diagram mengubah foto profil (anggota)

B.13. Keluar Aplikasi



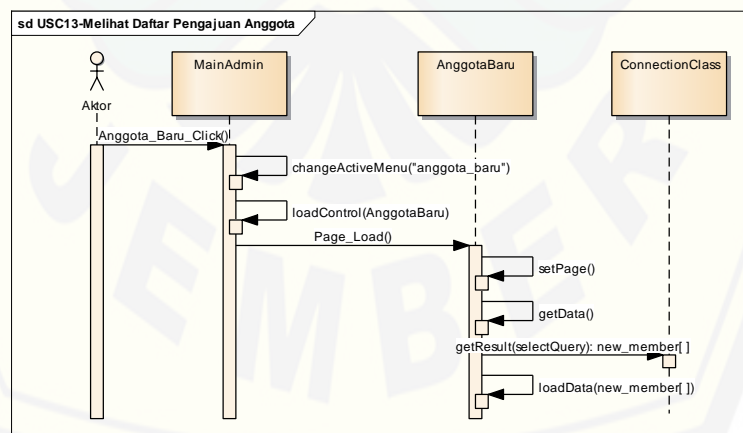
Gambar B. 14 Sequence diagram keluar aplikasi

B.14. Melihat Daftar Pengajuan Anggota



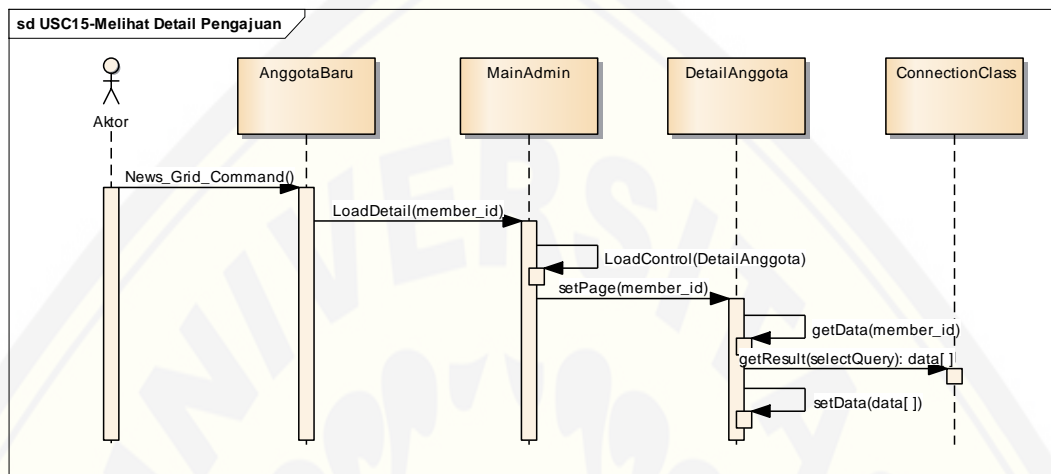
Gambar B. 15 Sequence diagram melihat daftar anggota

B.15. Mencari Daftar Pengajuan



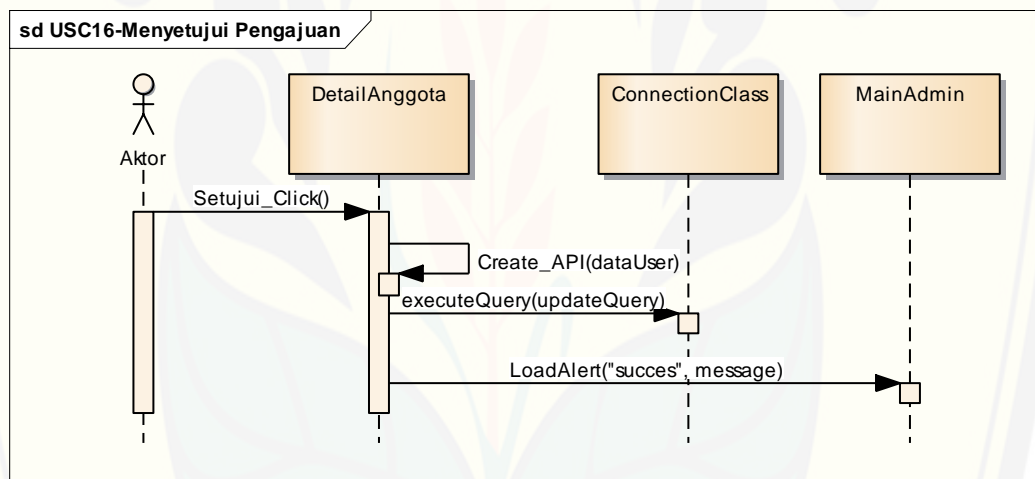
Gambar B. 16 Sequence diagram mencari daftar pengajuan

B.16. Melihat Detail Pengajuan



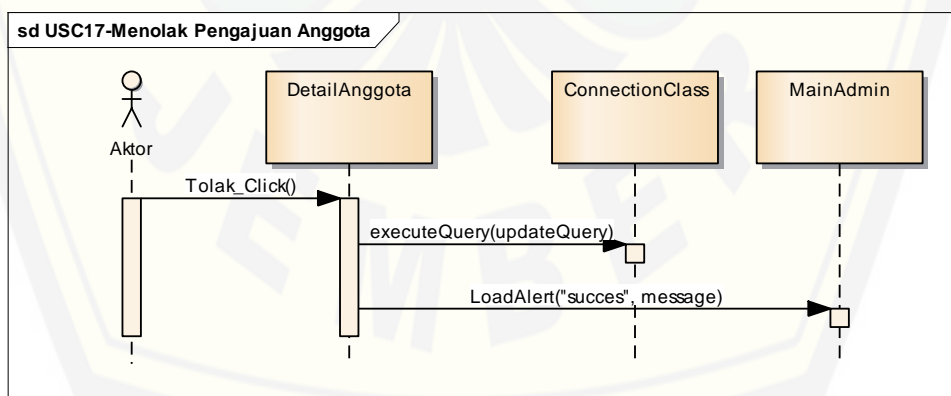
Gambar B. 17 Sequence diagram melihat detail pengajuan

B.17. Menyetujui Pengajuan Anggota



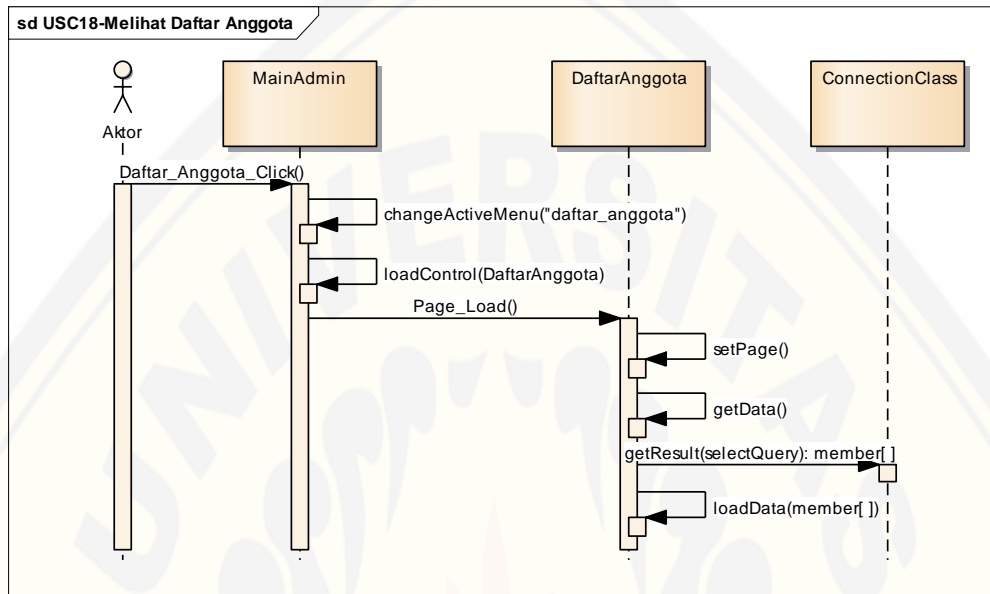
Gambar B. 18 Sequence diagram menyetujui pengajuan anggota

B.18. Menolak Pengajuan Anggota



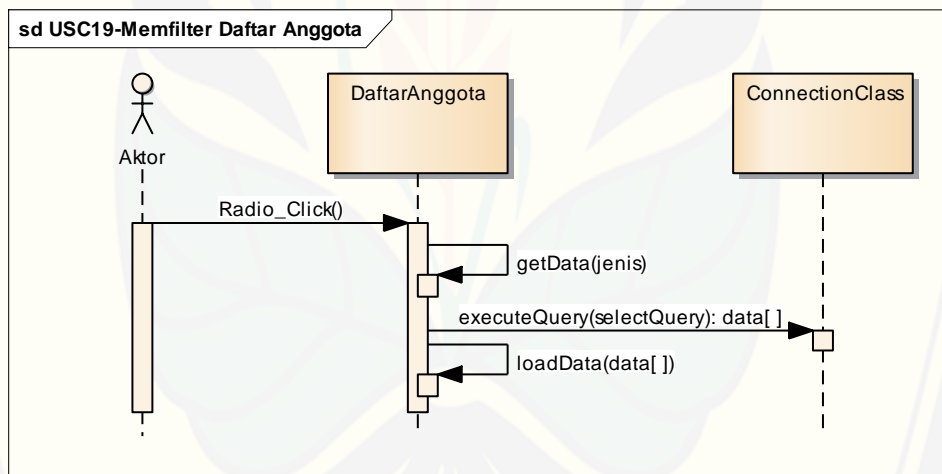
Gambar B. 19 Sequence diagram menolak pengajuan anggota

B.19. Melihat Daftar Anggota



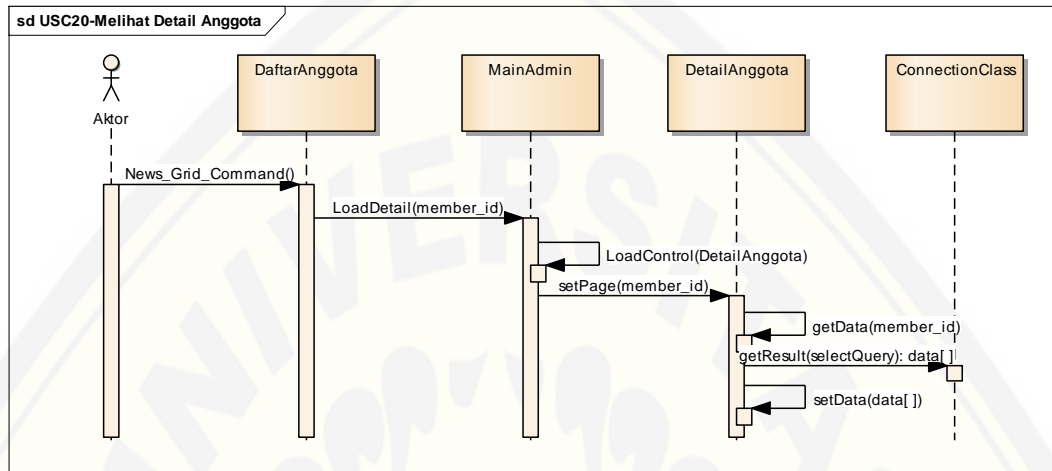
Gambar B. 20 Sequence diagram melihat daftar anggota

B.20. Memfilter Daftar Anggota



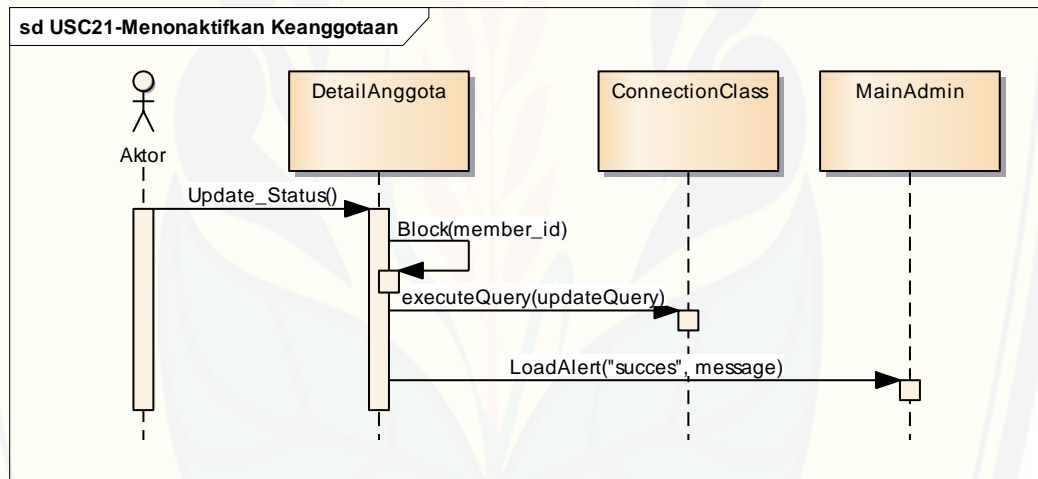
Gambar B. 21 Sequence diagram memfilter daftar anggota

B.21. Melihat Detail Anggota



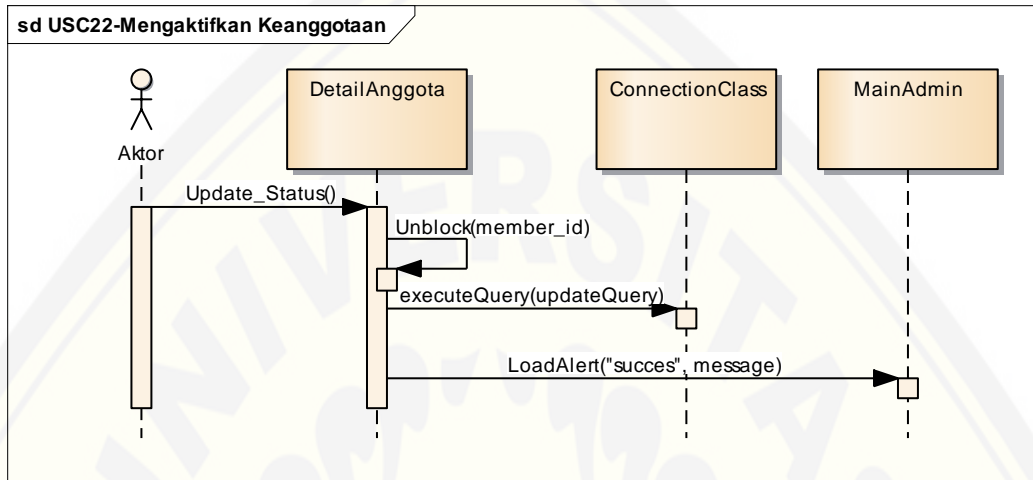
Gambar B. 22 Sequence diagram melihat detail anggota

B.22. Menonaktifkan Keanggotaan



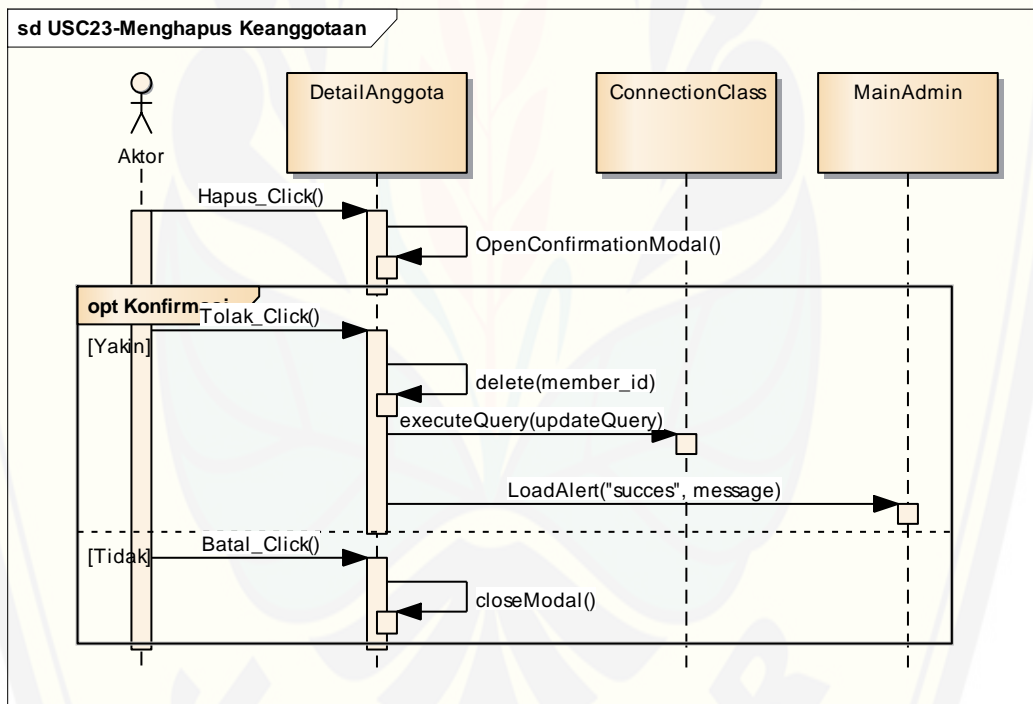
Gambar B. 23 Sequence diagram menonaktifkan keanggotaan

B.23. Mengaktifkan Keanggotaan



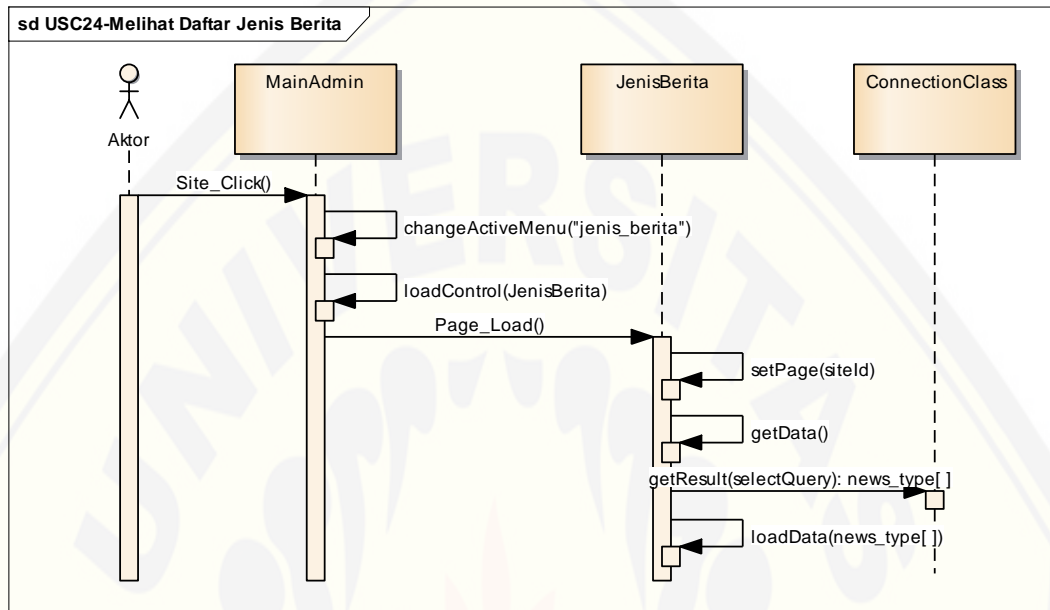
Gambar B. 24 Sequence diagram mengaktifkan keanggotaan

B.24. Menghapus Keanggotaan



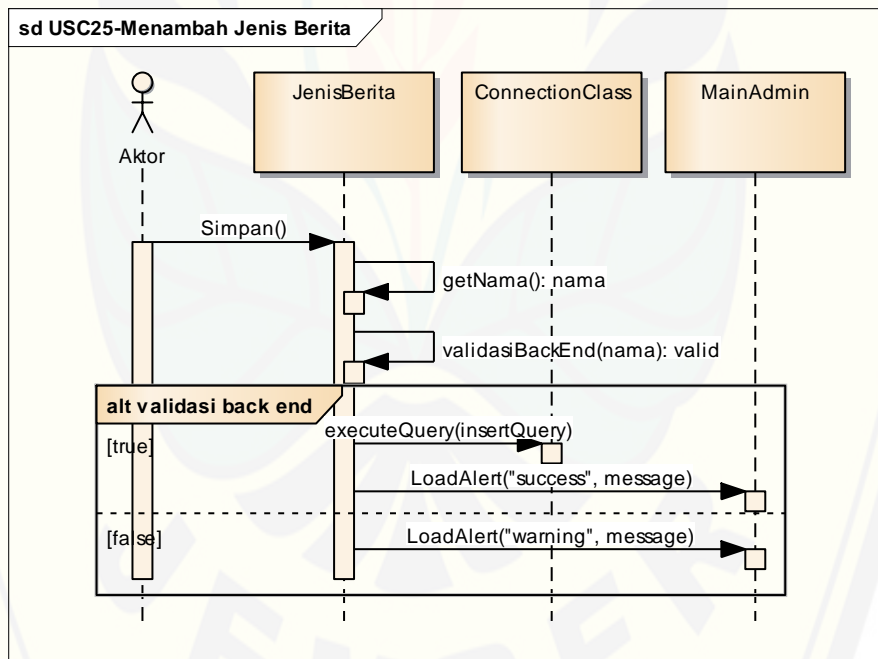
Gambar B. 25 Sequence diagram menghapus keanggotaan

B.25. Melihat Daftar Jenis Berita



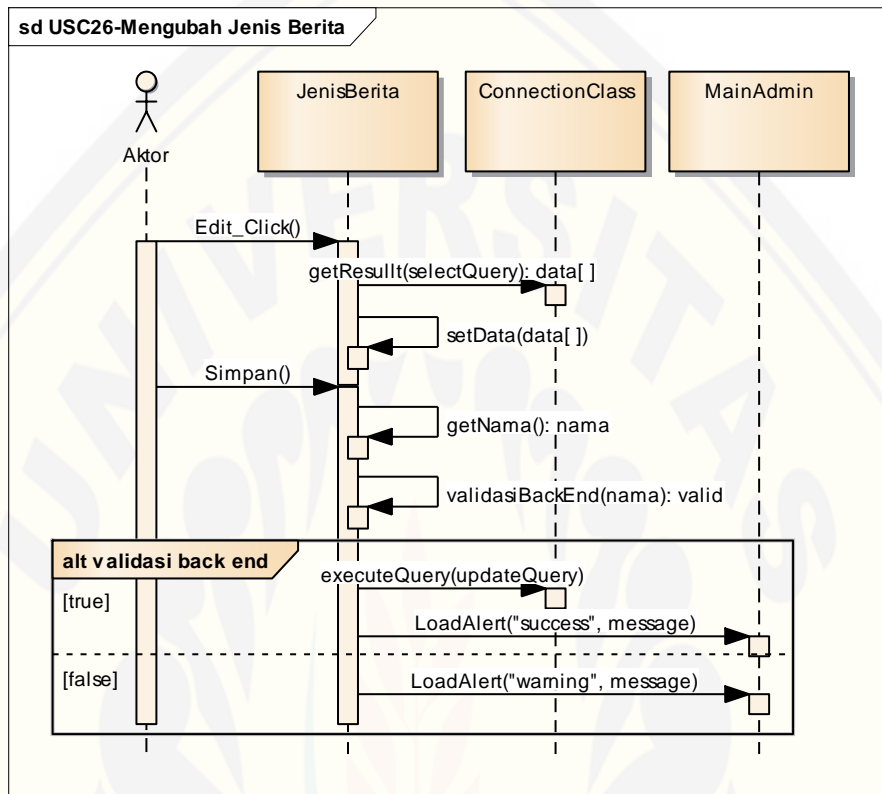
Gambar B. 26 Sequence diagram melihat jenis berita

B.26. Menambahkan Jenis Berita Baru



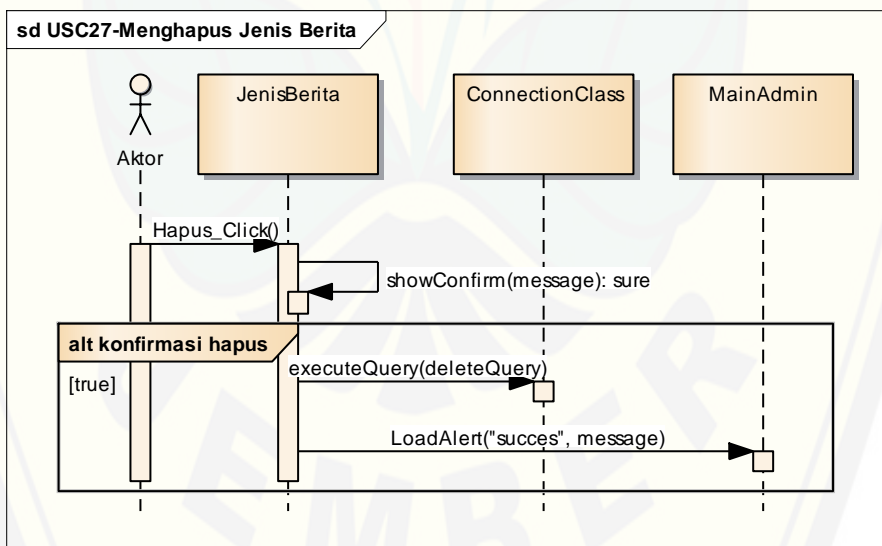
Gambar B. 27 Sequence diagram menambahkan jenis berita

B.27. Mengubah Jenis Berita



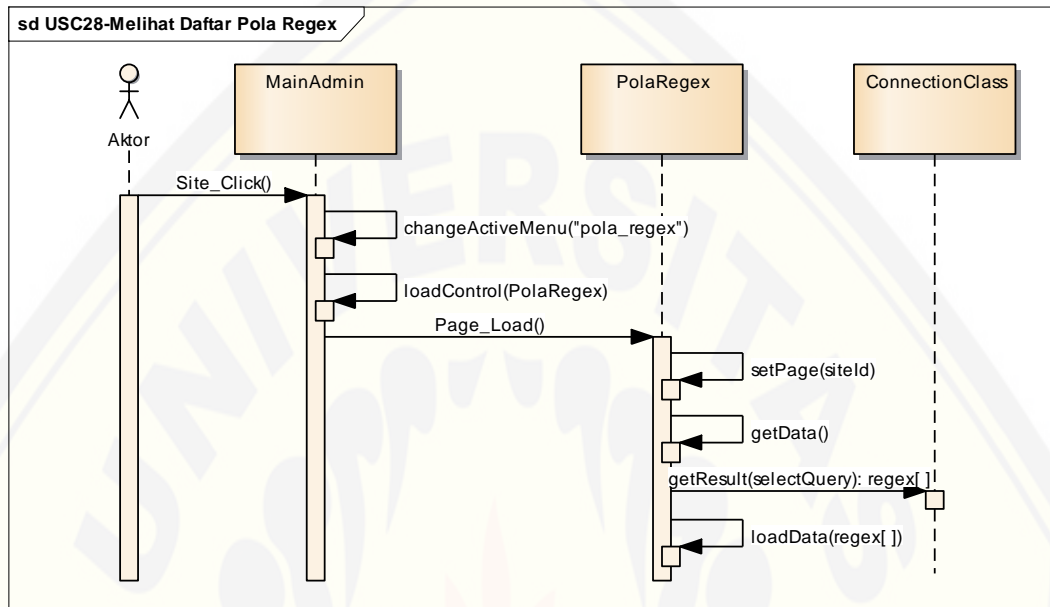
Gambar B. 28 Sequence diagram mengubah jenis berita

B.28. Menghapus Jenis Berita



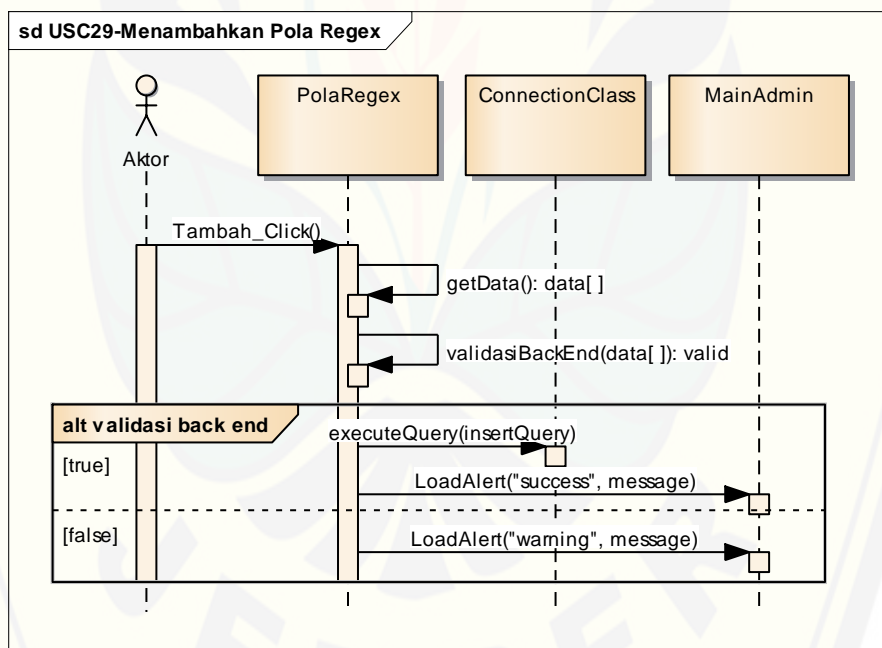
Gambar B. 29 Sequence diagram menghapus jenis berita

B.29. Melihat Daftar Pola Ekspresi Regular



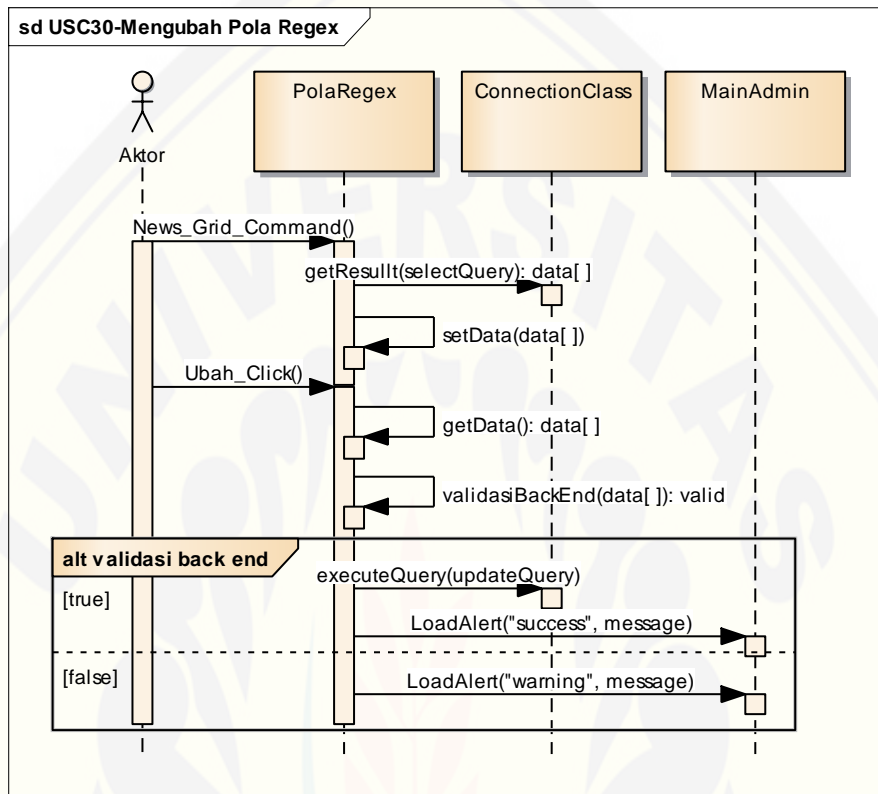
Gambar B. 30 Sequence diagram melihat daftar pola regex

B.30. Menambahkan Pola Eskpresi Regular



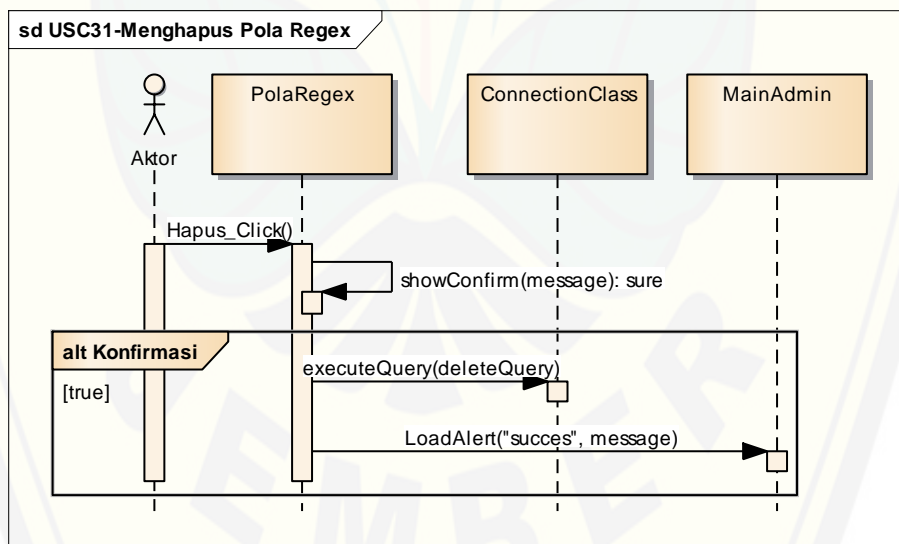
Gambar B. 31 Sequence diagram menambahkan pola regex

B.31. Mengubah Pola Ekspresi Regular



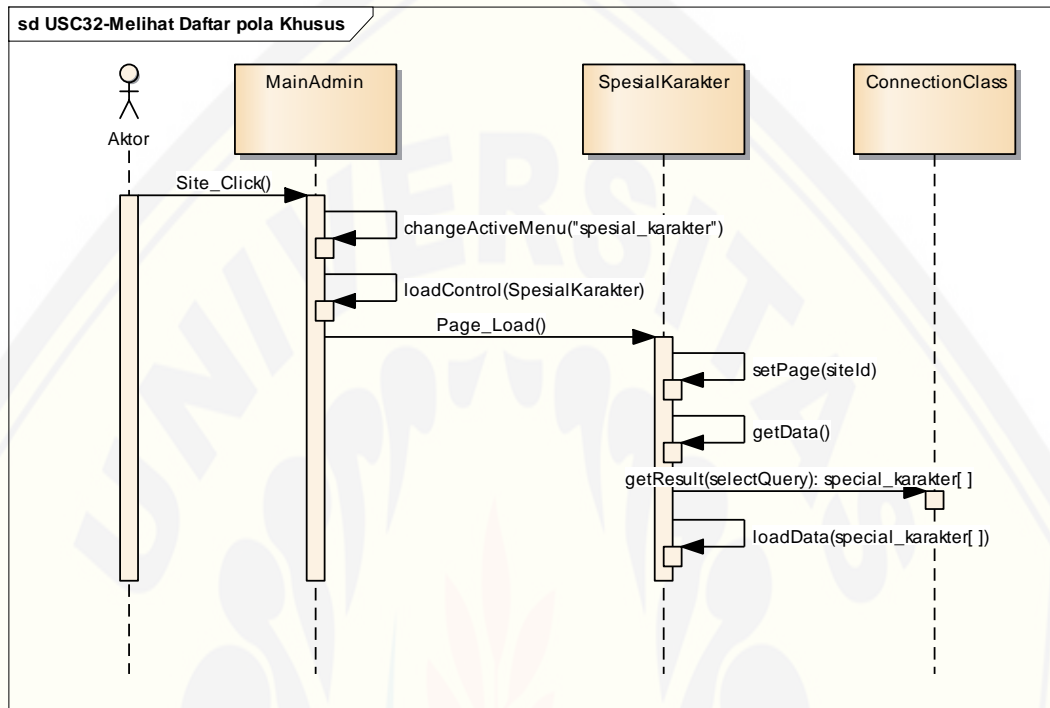
Gambar B. 32 Sequence diagram mengubah pola regex

B.32. Menghapus Pola Ekspresi Regular



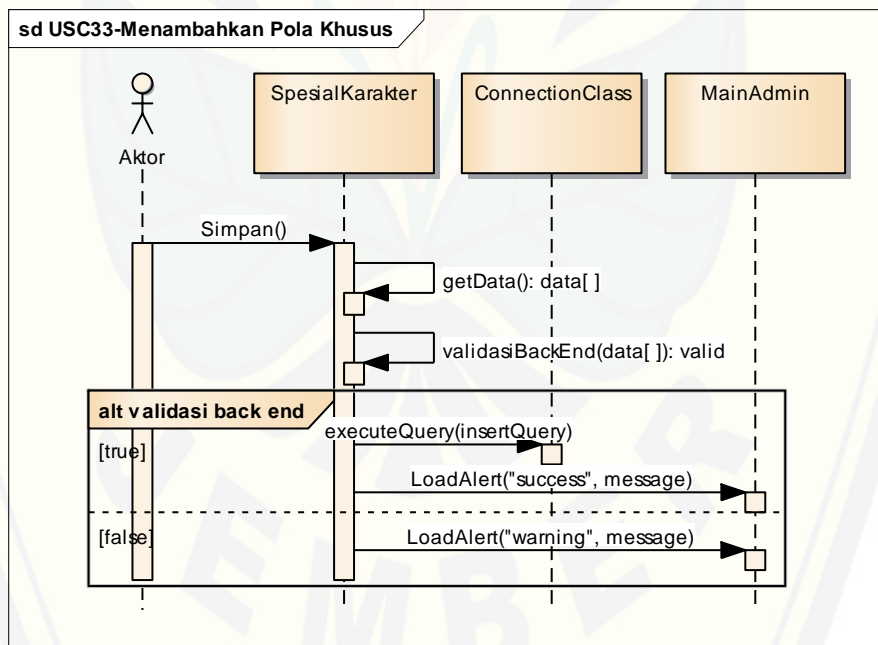
Gambar B. 33 Sequence diagram menghapus pola regex

B.33. Melihat Daftar Pola Khusus



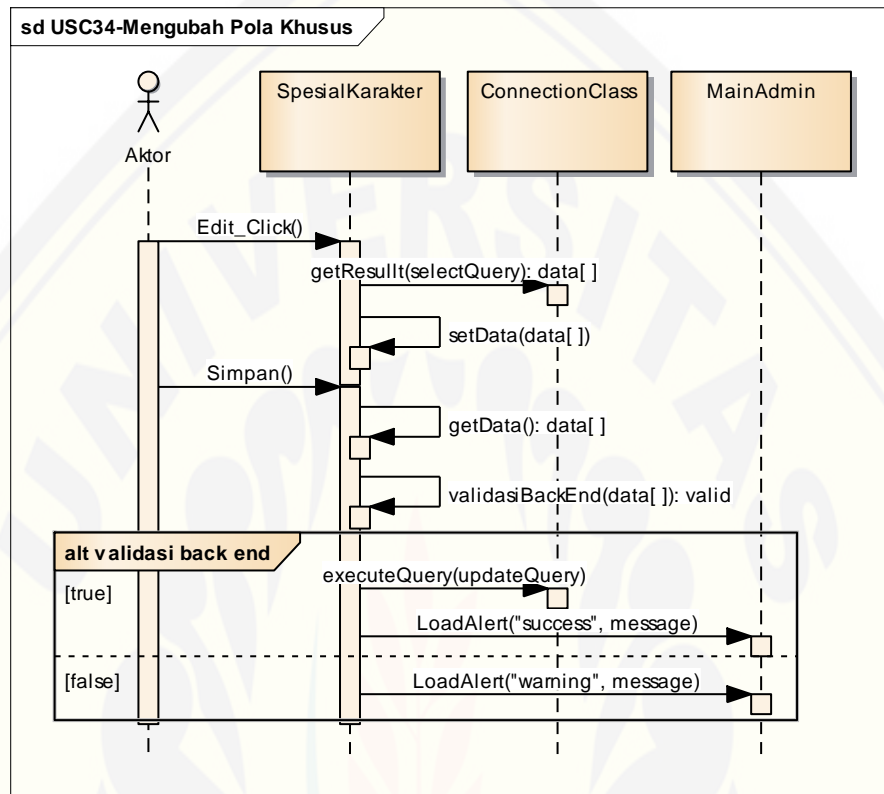
Gambar B. 34 Sequence diagram melihat daftar pola khusus

B.34. Menambahkan Daftar Pola Khusus



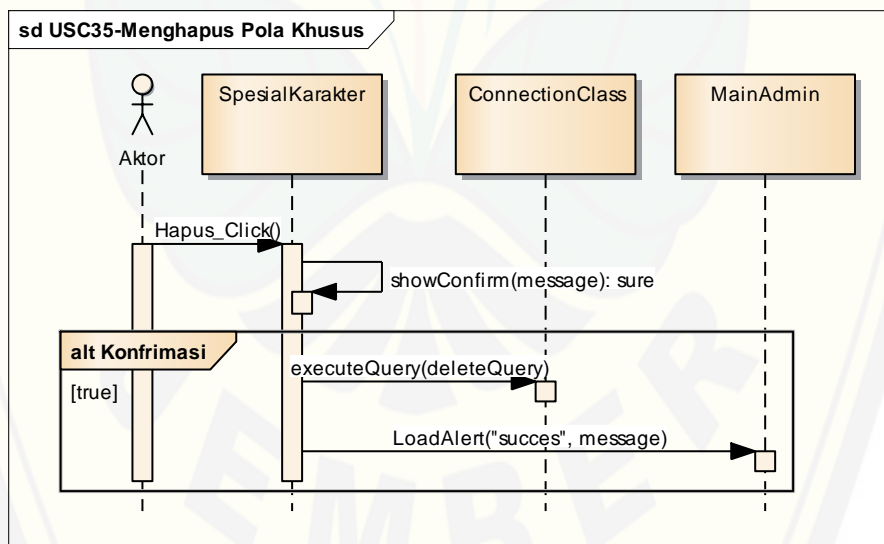
Gambar B. 35 Sequence diagram menambahkan pola khusus

B.35. Mengubah Pola Khusus



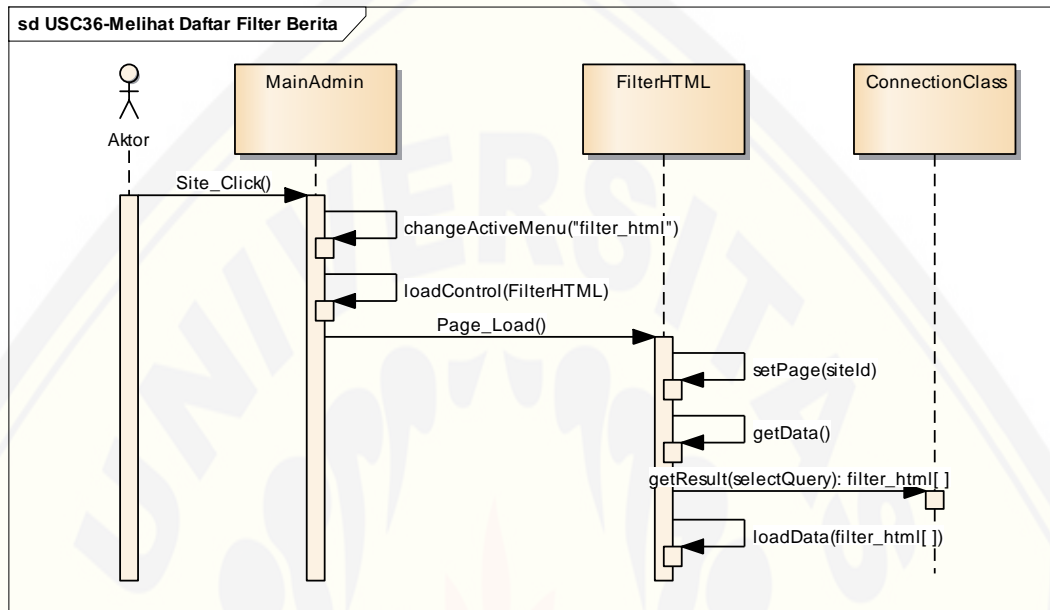
Gambar B. 36 Sequence diagram mengubah pola khusus

B.36. Menghapus Pola Khusus



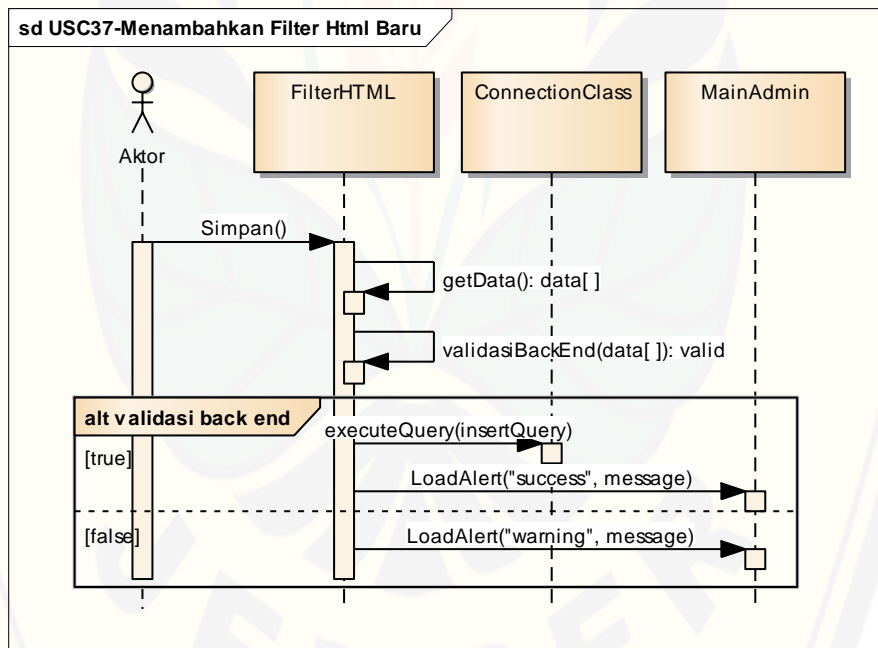
Gambar B. 37 Sequence diagram menghapus pola khusus

B.37. Melihat Daftar Filter HTML



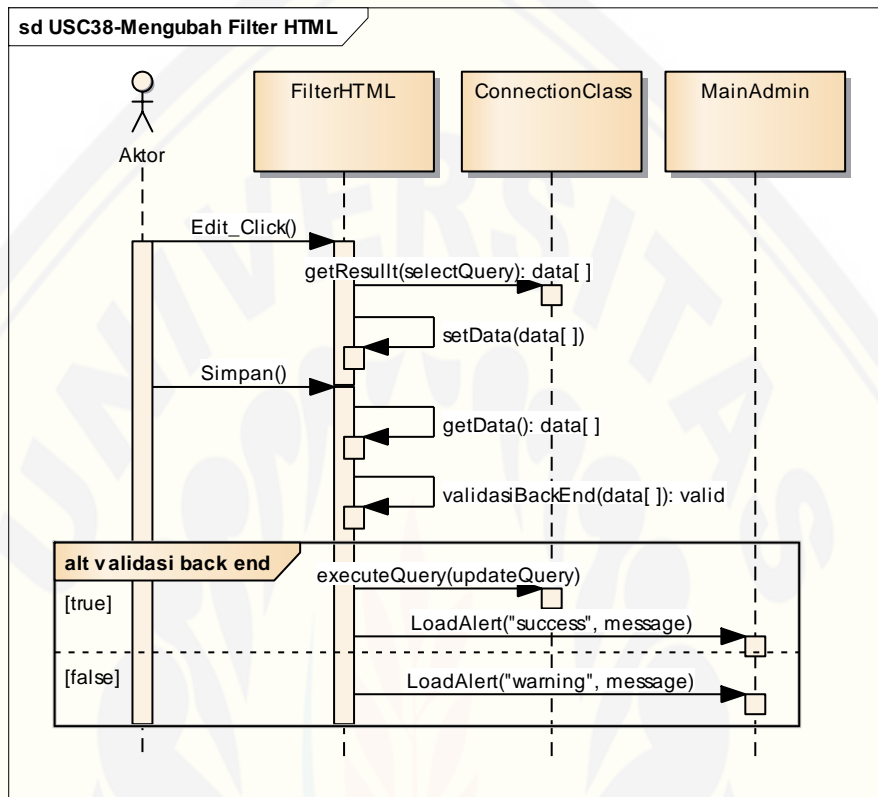
Gambar B. 38 Sequence diagram melihat filter html

B.38. Menambahkan Filter HTML Baru



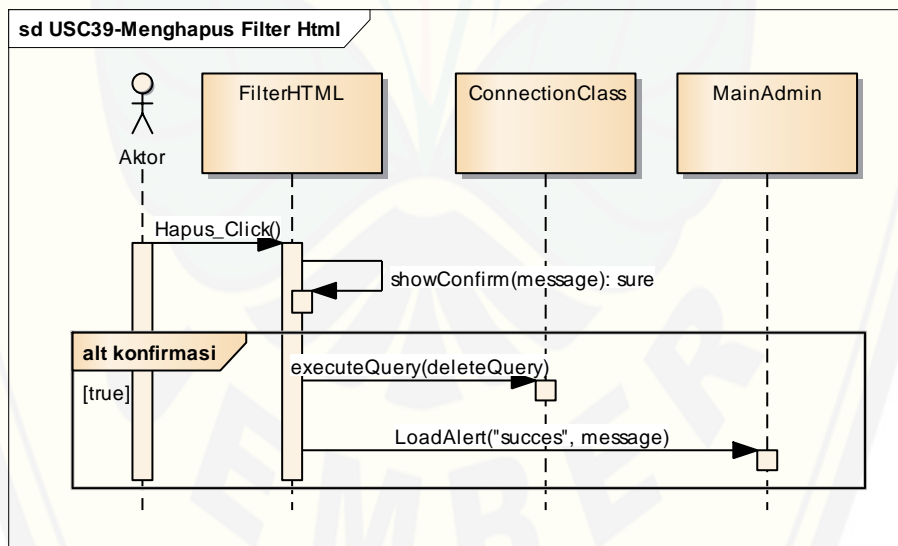
Gambar B. 39 Sequence diagram menambahkan filter html

B.39. Mengubah Filter HTML



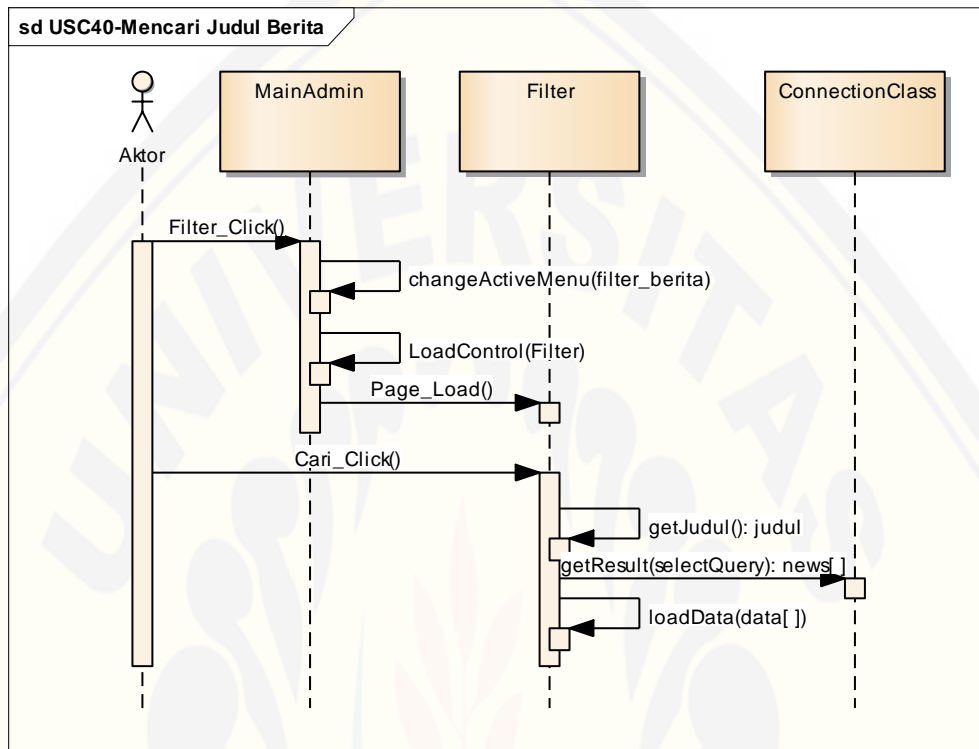
Gambar B. 40 Sequence diagram mengubah filter html

B.40. Menghapus Filter HTML



Gambar B. 41 Sequence diagram menghapus filter html

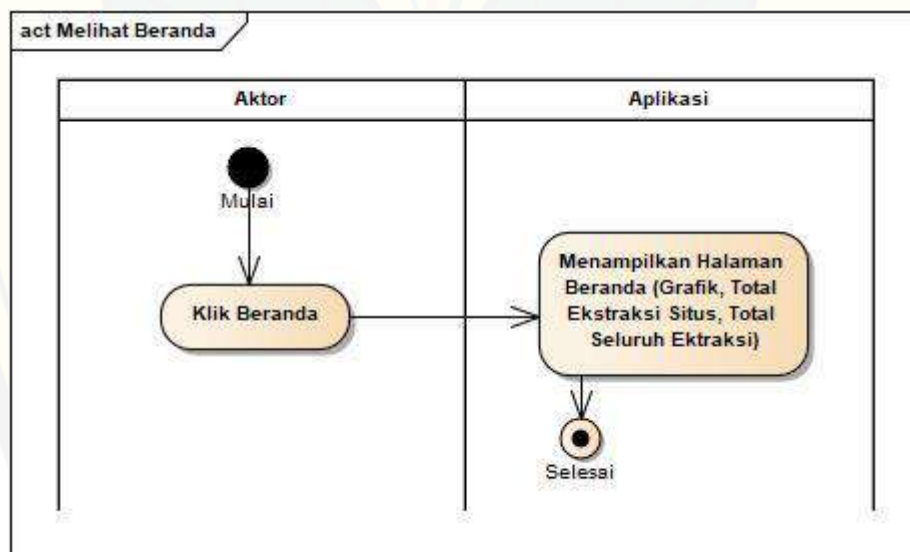
B.41. Mencari Judul Berita



Gambar B. 42 Sequence diagram mencari judul berita

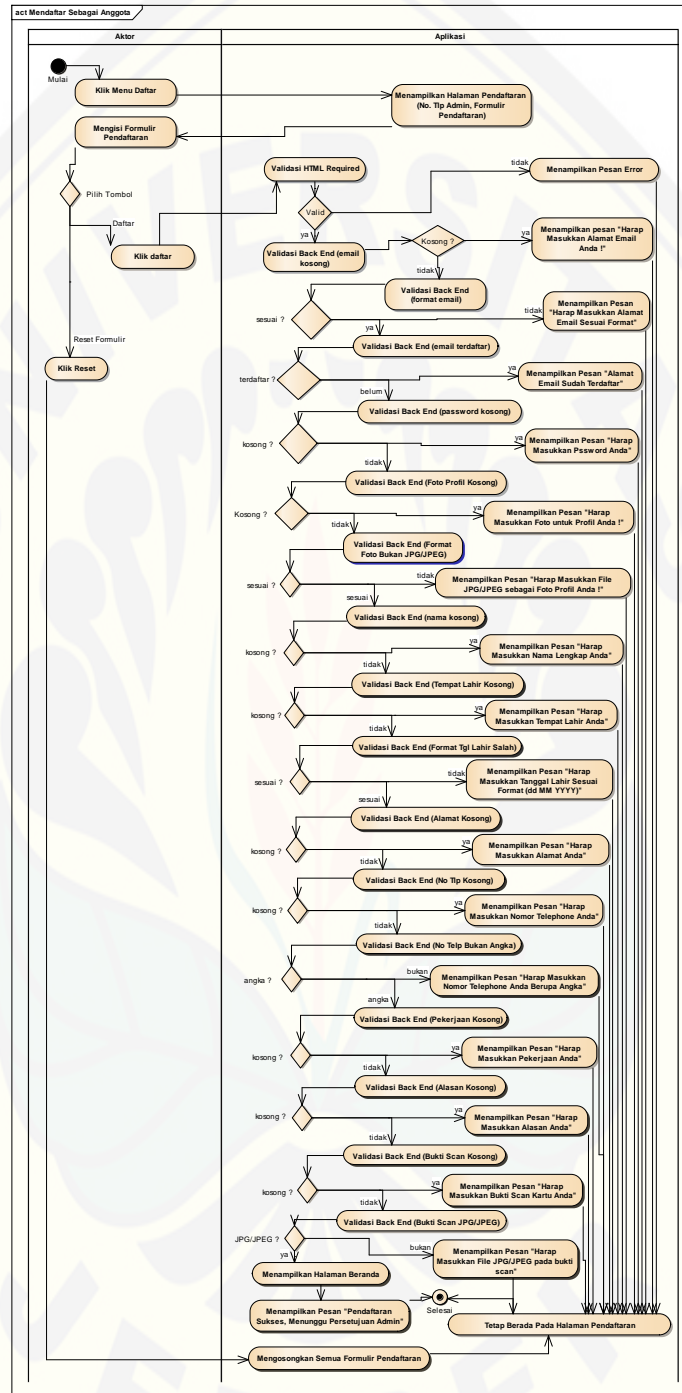
Lampiran C. Activity Diagram

C.1. Melihat Beranda



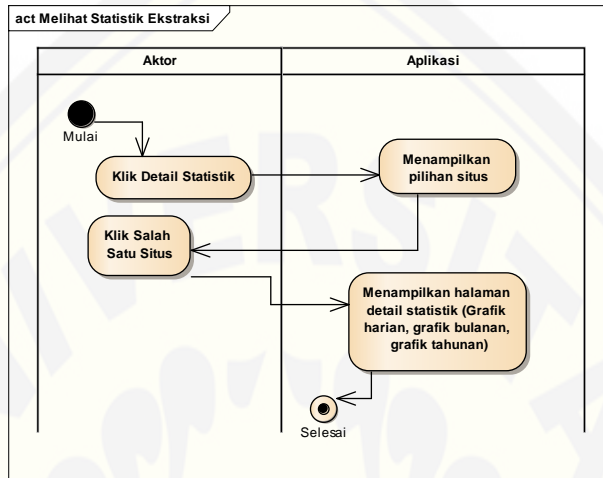
Gambar C. 1 Activity Diagram Melihat Beranda

C.2. Mendaftar sebagai Anggota



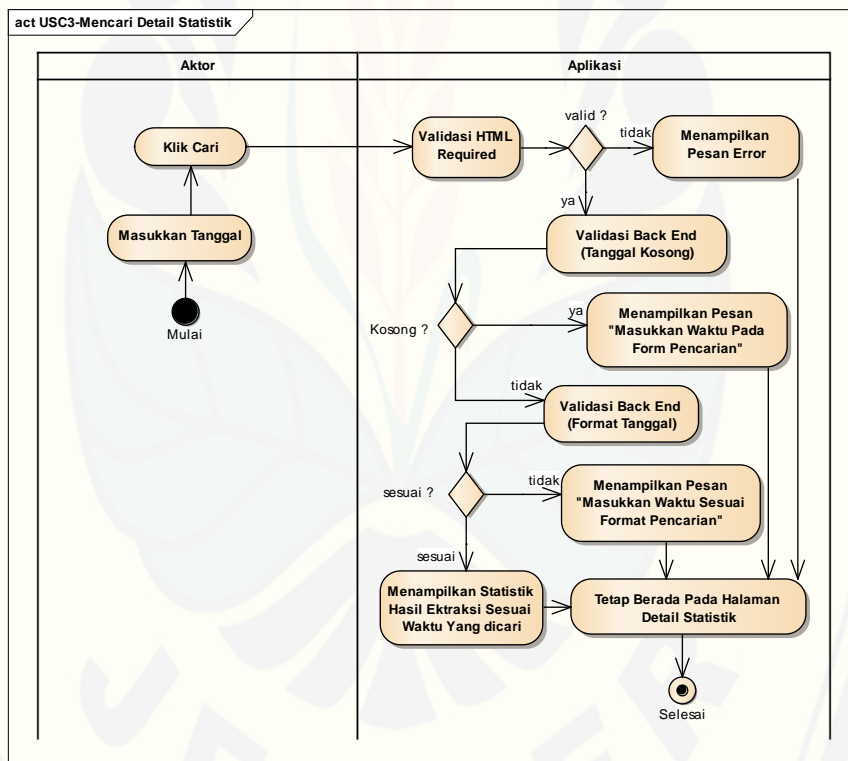
Gambar C. 2 Activity Diagram Mendaftar Anggota

C.3. Melihat Statistik Ekstraksi



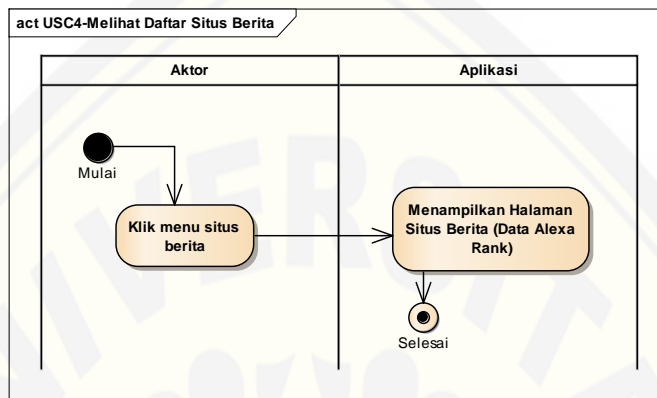
Gambar C. 3 Activity Diagram Melihat Statistik

C.4. Mencari Detail Statistik



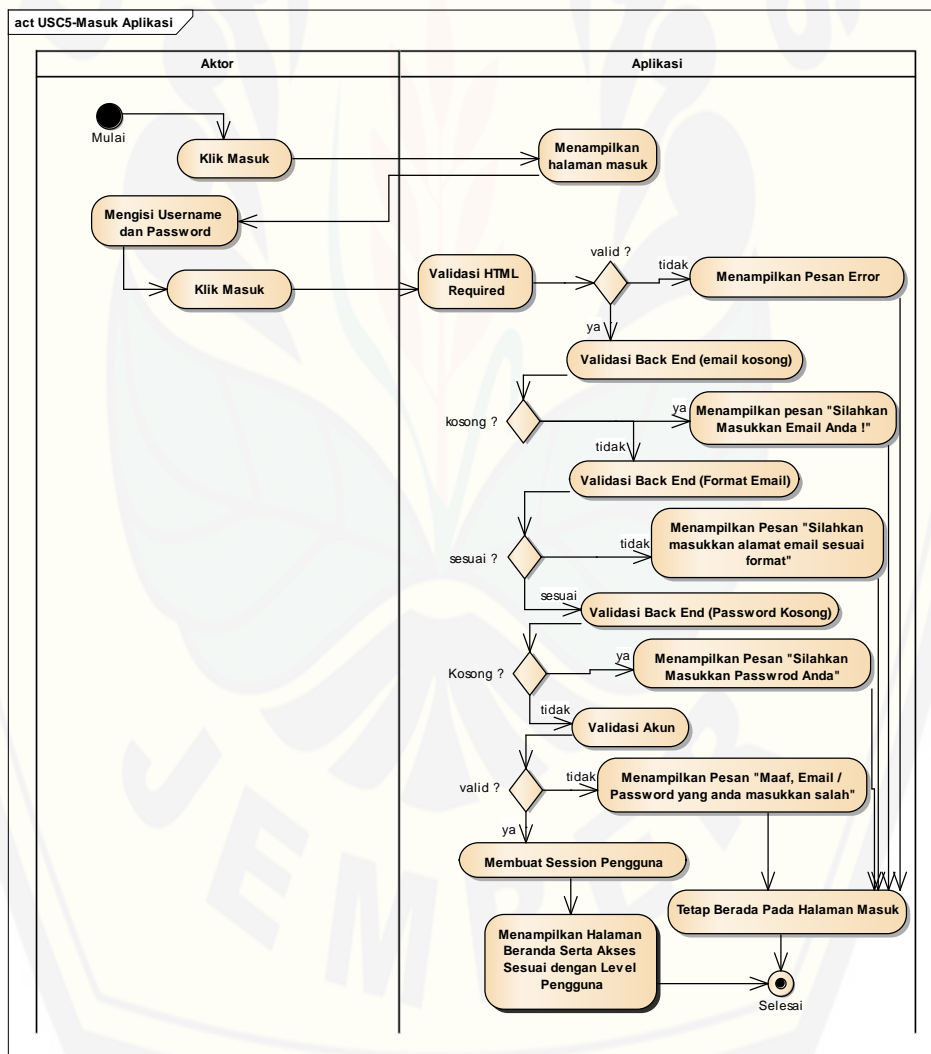
Gambar C. 4 Activity Diagram Mencari Detail Statistik

C.5. Melihat Daftar Situs Berita



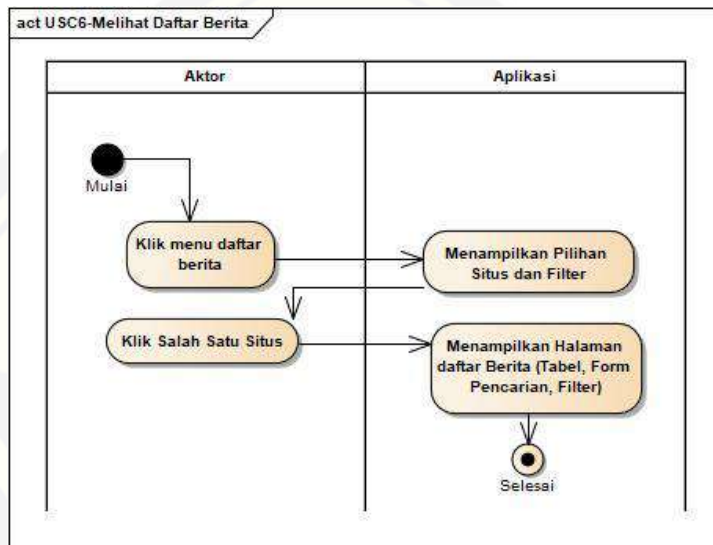
Gambar C. 5 Activity Diagram Melihat Daftar Situs Berita

C.6. Masuk Aplikasi



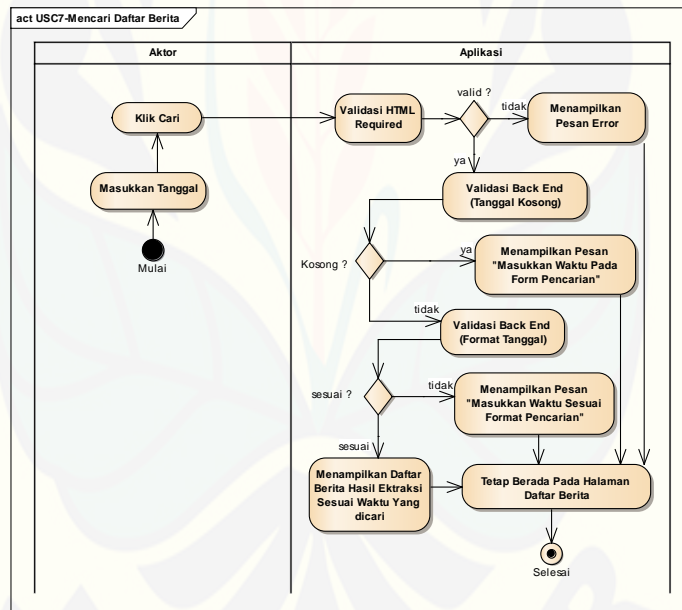
Gambar C. 6 Activity Diagram Masuk Aplikasi

C.7. Melihat Daftar Berita



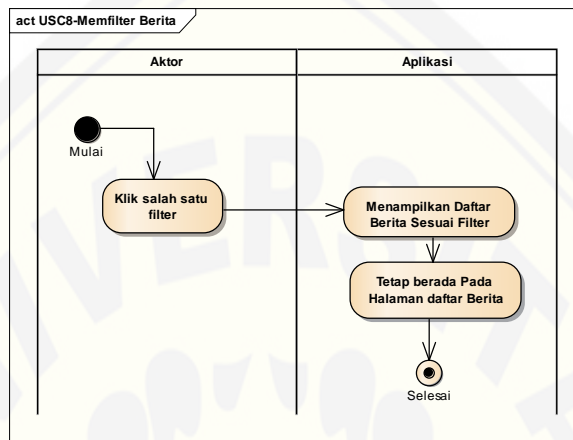
Gambar C. 7 Activity Diagram Melihat Daftar Berita

C.8. Mencari Daftar Berita



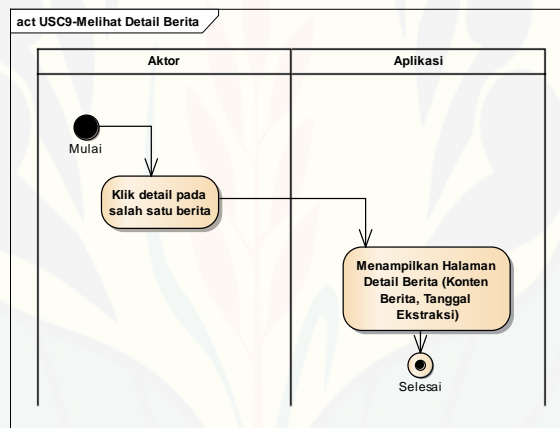
Gambar C. 8 Activity Diagram Mencari Daftar Berita

C.9. Memfilter Daftar Berita



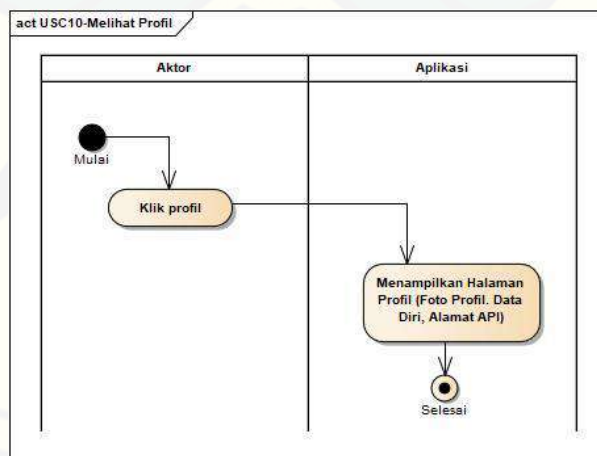
Gambar C. 9 Activity Diagram Memfilter Daftar Berita

C.10. Melihat Detail Berita



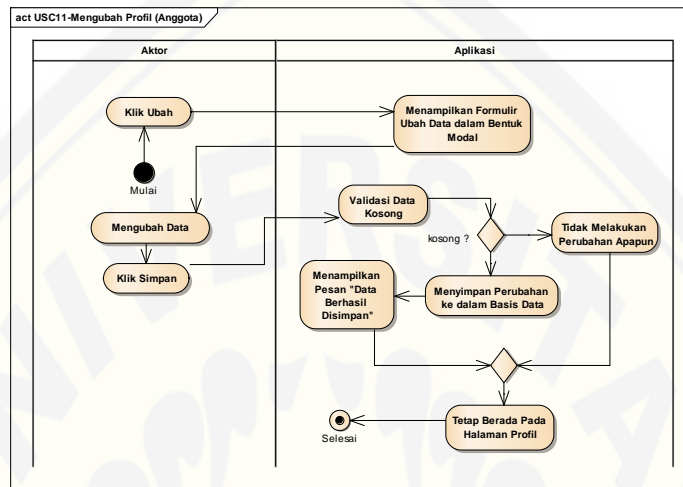
Gambar C. 10 Activity Diagram Melihat Detail Berita

C.11. Melihat Profil

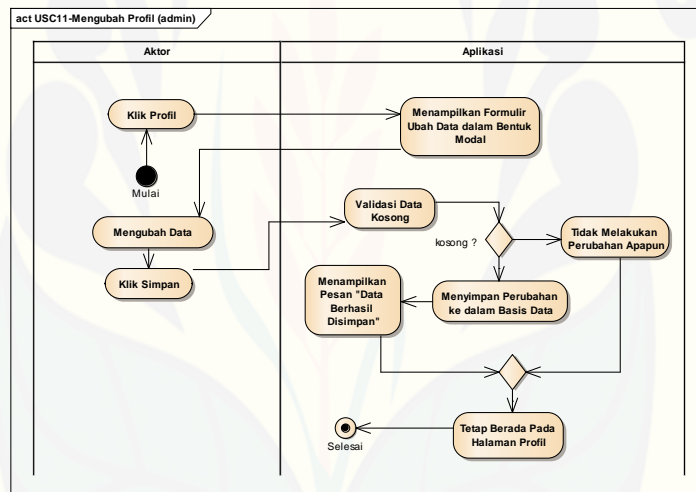


Gambar C. 11 Activity Diagram Melihat Profil

C.12. Mengubah Profil

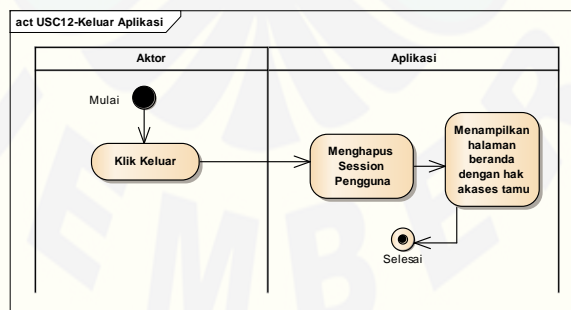


Gambar C. 12 Activity Diagram Mengubah Profil (admin)



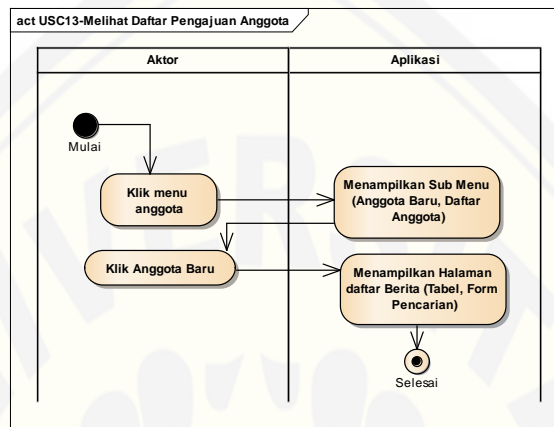
Gambar C. 13 Activity Diagram Mengubah Profil (anggota)

C.13. Keluar Aplikasi



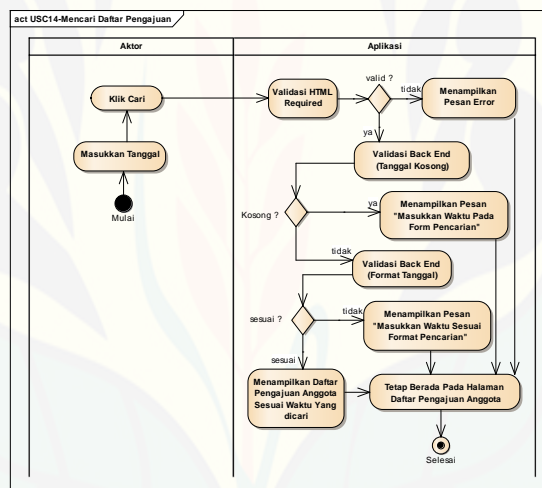
Gambar C. 14 Activity Diagram Keluar Aplikasi

C.14. Melihat Daftar Pengajuan Anggota



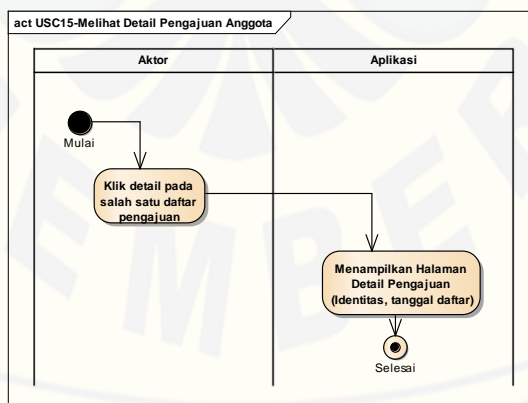
Gambar C. 15 Activity Diagram Melihat Daftar Pengajuan Anggota

C.15. Mencari Daftar Pengajuan



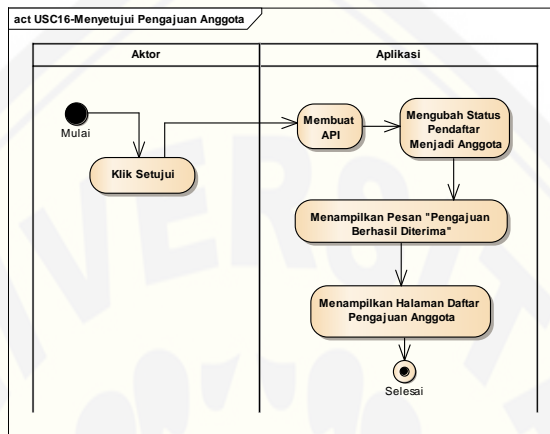
Gambar C. 16 Activity Diagram Mencari Daftar Pengajuan

C.16. Melihat Detail Pengajuan



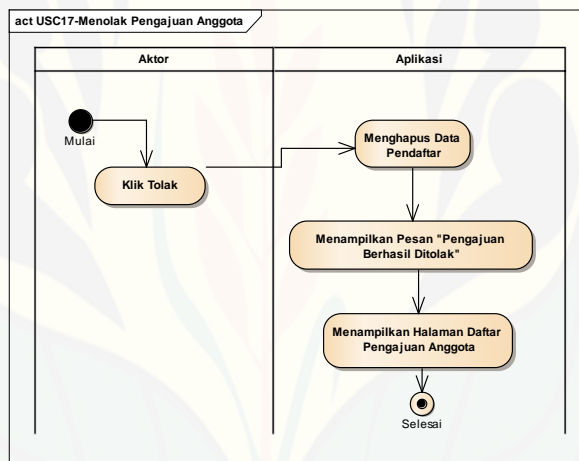
Gambar C. 17 Activity Diagram Melihat Detail Pengajuan

C.17. Menyetujui Pengajuan Anggota



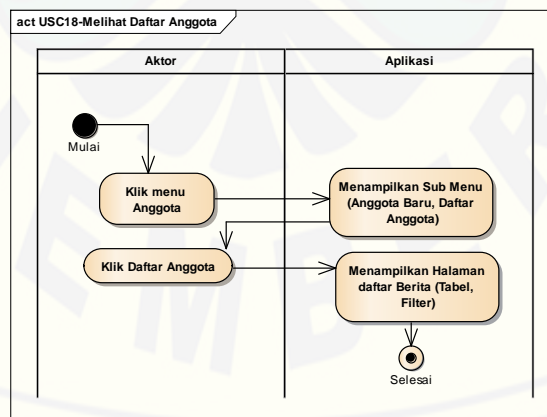
Gambar C. 18 Activity Diagram Menyetujui Pengajuan

C.18. Menolak Pengajuan Anggota



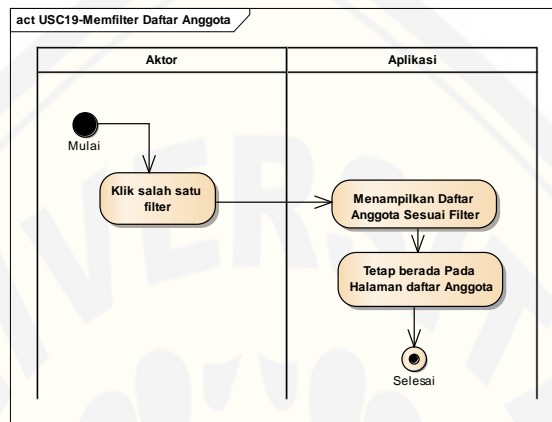
Gambar C. 19 Activity Diagram Menolak Pengajuan Anggota

C.19. Melihat Daftar Anggota



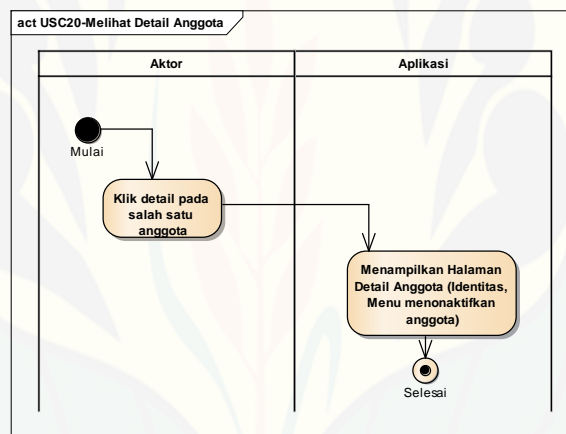
Gambar C. 20 Activity Diagram Melihat Daftar Anggota

C.20. Memfilter Daftar Anggota



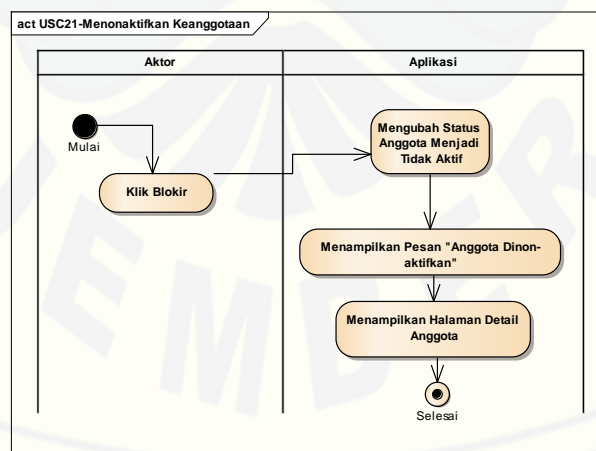
Gambar C. 21 Activity Diagram Memfilter Daftar Anggota

C.21. Melihat Detail Anggota



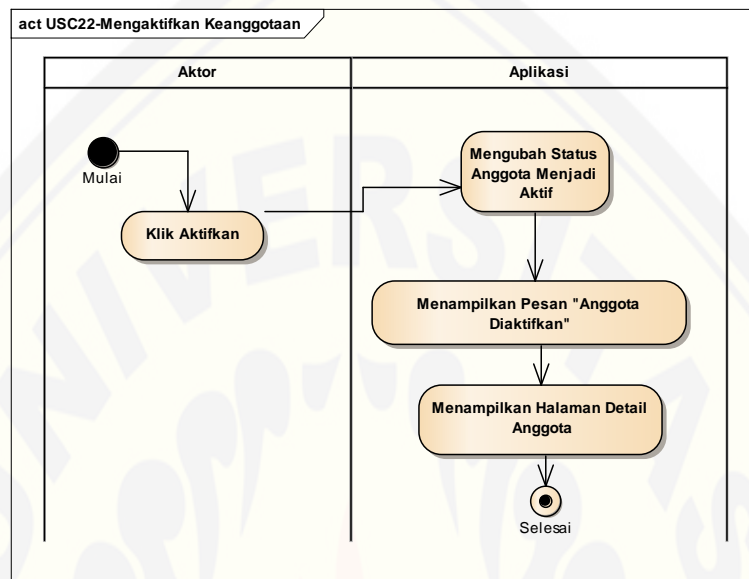
Gambar C. 22 Activity Diagram Melihat Detail Anggota

C.22. Menonaktifkan Keanggotaan



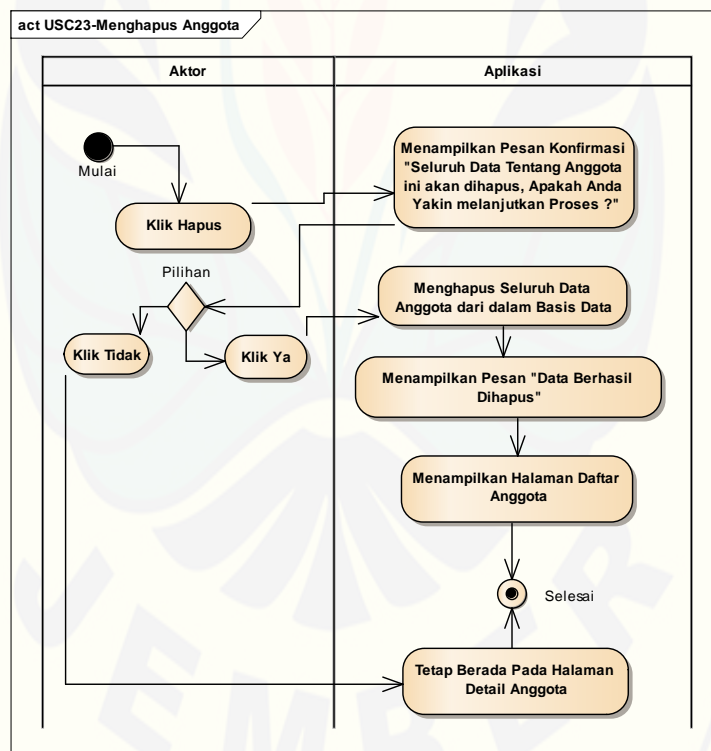
Gambar C. 23 Activity Diagram Menonaktifkan Keanggotaan

C.23. Mengaktifkan Keanggotaan



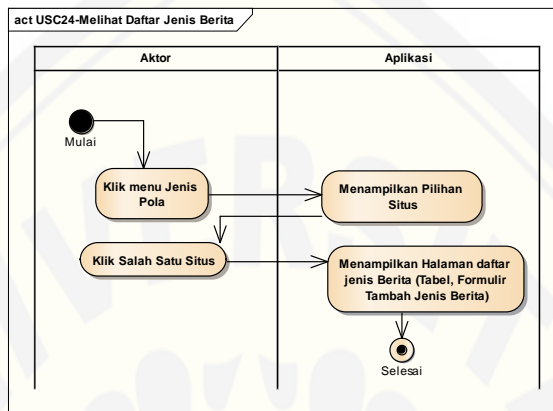
Gambar C. 24 Activity Diagram Mengaktifkan Keanggotaan

C.24. Menghapus Keanggotaan



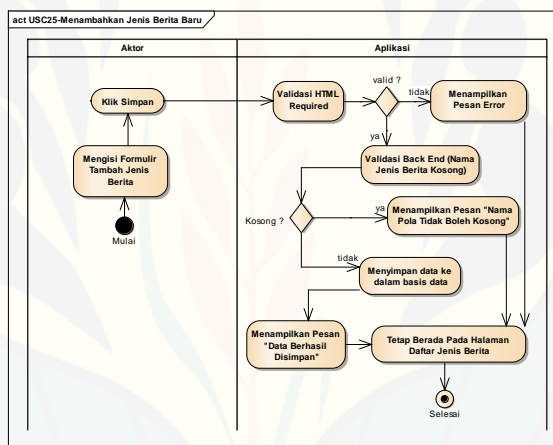
Gambar C. 25 Activity Diagram Menghapus Keanggotaan

C.25. Melihat Daftar Jenis Berita



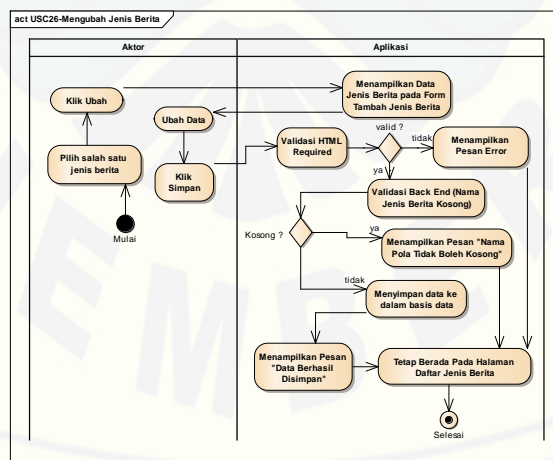
Gambar C. 26 Activity Diagram Melihat Daftar Jenis Berita

C.26. Menambahkan Jenis Berita Baru



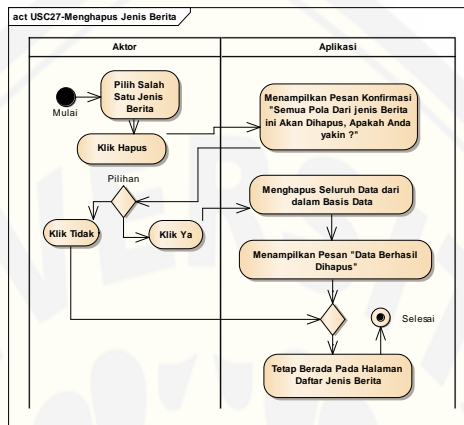
Gambar C. 27 Activity Diagram Menambah Jenis Berita

C.27. Mengubah Jenis Berita



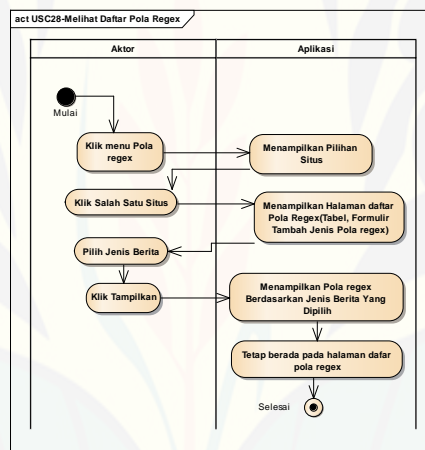
Gambar C. 28 Activity Diagram Mengubah Jenis Berita

C.28. Menghapus Jenis Berita



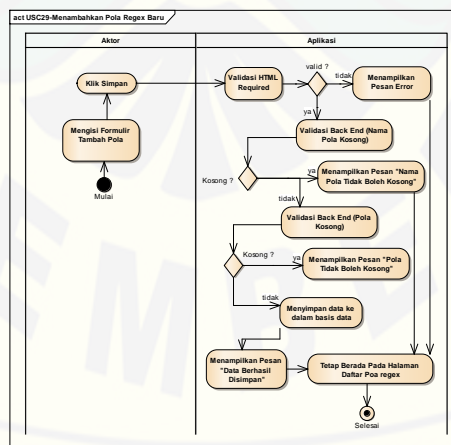
Gambar C. 29 Activity Diagram Menghapus Jenis Berita

C.29. Melihat Daftar Pola Ekspresi Regular



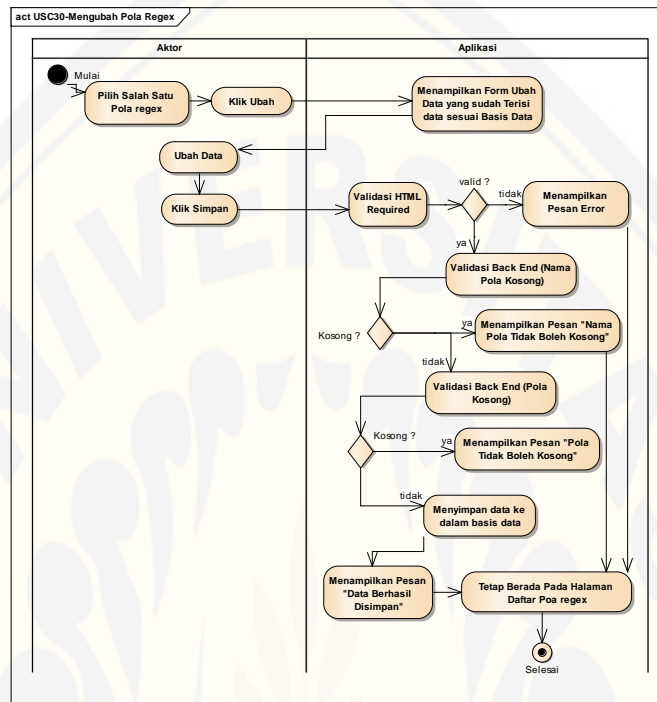
Gambar C. 30 Activity Diagram Melihat Pola Regex

C.30. Menambahkan Pola Ekspresi Regular



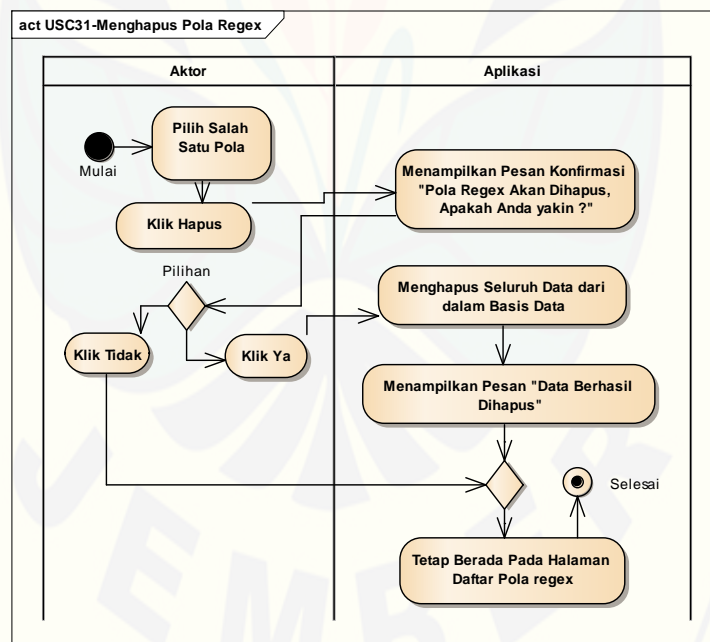
Gambar C. 31 Activity Diagram Menambah Pola Regex

C.31. Mengubah Pola Ekspresi Regular



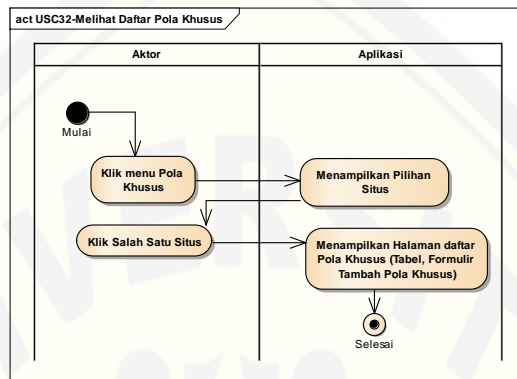
Gambar C. 32 Activity Diagram Mengubah Pola Regex

C.32. Menghapus Pola Ekspresi Reular



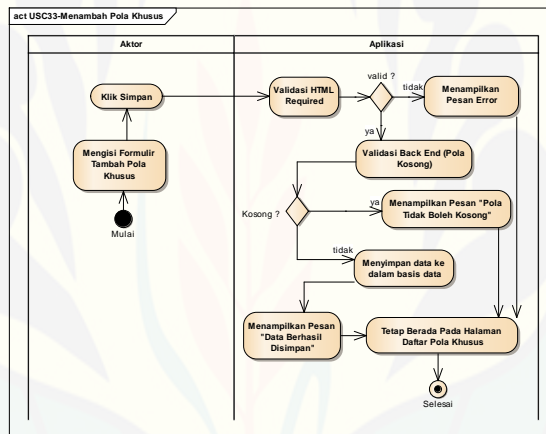
Gambar C. 33 Activity Diagram Menghapus Pola Regex

C.33. Melihat Daftar Pola Khusus



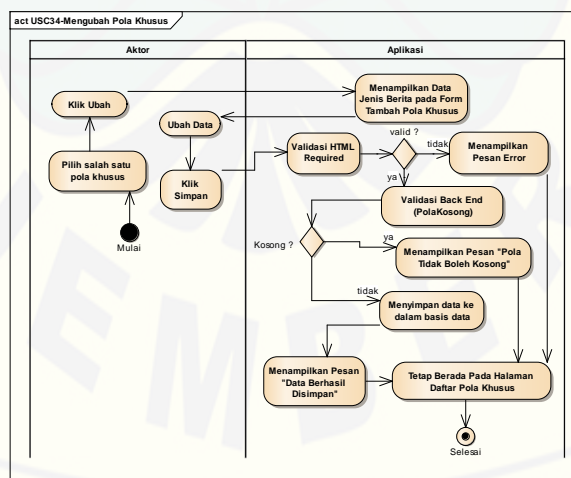
Gambar C. 34 Activity Diagram Melihat Daftar Pola Khusus

C.34. Menambahkan Daftar Pola Khusus



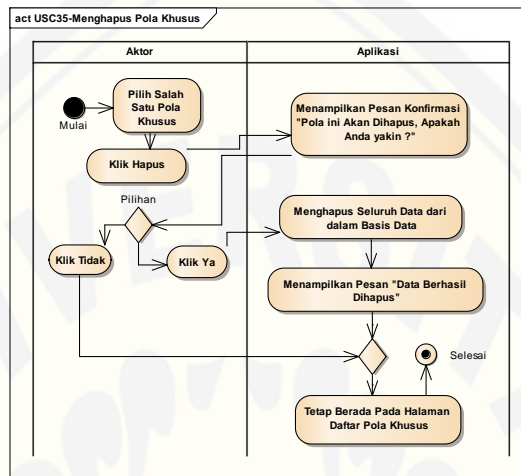
Gambar C. 35 Activity Diagram Menambah Pola Khusus

C.35. Mengubah Pola Khusus



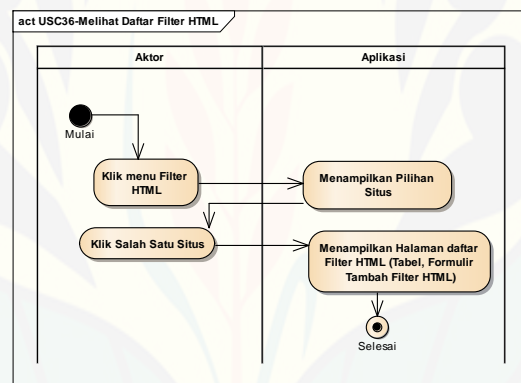
Gambar C. 36 Activity Diagram Mengubah Pola Khusus

C.36. Menghapus Pola Khusus



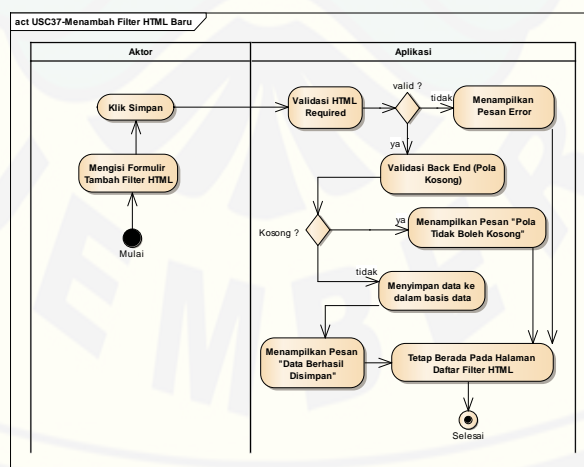
Gambar C. 37 Activity Diagram Menghapus Pola Khusus

C.37. Melihat Daftar Filter HTML



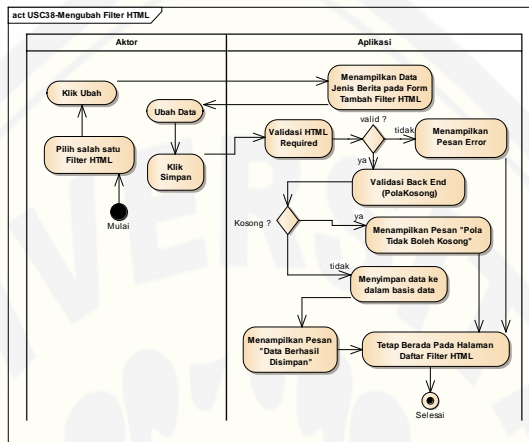
Gambar C. 38 Activity Diagram Melihat Daftar Filter HTML

C.38. Menambahkan Filter HTML Baru



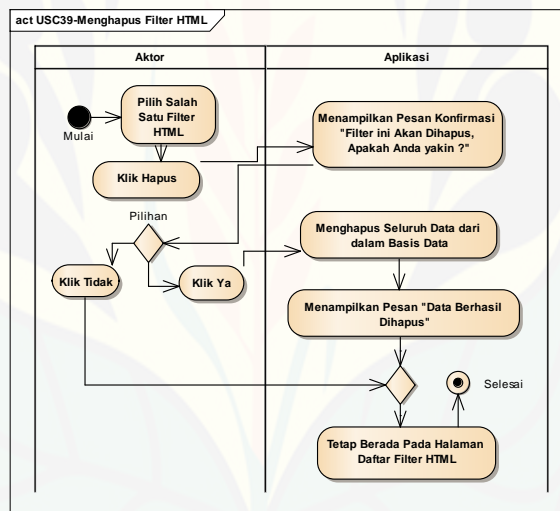
Gambar C. 39 Activity Diagram Menambah Filter HTML

C.39. Mengubah Filter HTML



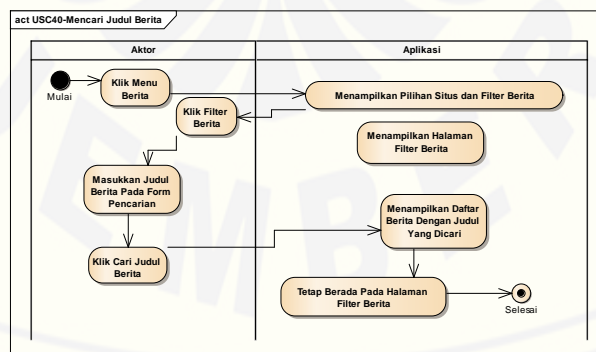
Gambar C. 40 Activity Diagram Mengubah Filter HTML

C.40. Menghapus Filter HTML



Gambar C. 41 Menghapus Filter HTML Menghapus Filter HTML

C.41. Mencari Judul Berita



Gambar C. 42 Activity Diagram Mencari Judul Berita