

**PEMBANGUNAN PERANGKAT LUNAK WEB SCRAPING UNTUK
SITUS BERITA DAN PERINGKAS BERITA**

TUGAS AKHIR

**Diajukan Untuk Memenuhi Sebagian Persyaratan
Mencapai Derajat Sarjana Teknik Informatika**



Oleh:

YUSTINUS ANDRY SETIAWAN

11 07 06659

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2015**

HALAMAN PENGESAHAN

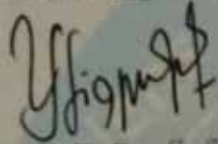
Tugas Akhir Berjudul

PEMBANGUNAN PERANGKAT LUNAK WEB SCRAPING UNTUK SITUS
BERITA DAN PERINGKAS BERITA

Disusun Oleh :
Yustinus Andry Setiawan (NIM : 11 07 06659)

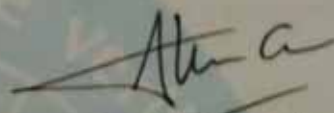
Dinyatakan telah memenuhi syarat
Pada tanggal : Oktober 2015

Pembimbing I



Y. Sigit Purnomo W.P, S.T., M.Kom.

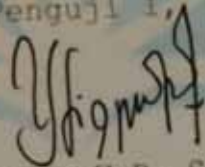
Pembimbing II



Kusworo Anindito, S.T., M.T.

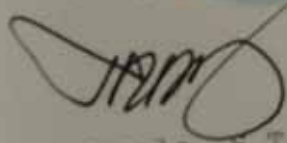
Tim Penguji :

Penguji I,



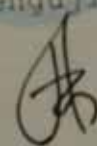
Y. Sigit Purnomo W.P, S.T., M.Kom.

Penguji II,



Thomas Suselo, S.T., M.T.

Penguji III,



Th. Adi Purnomo Sidhi, S.T., M.T.

Yogyakarta, Oktober 2015
Universitas Atma Jaya Yogyakarta
Fakultas Teknologi Industri



Dekan,
Dr. A. Teguh Siswanto

HALAMAN PERNYATAAN

Saya yang bertanda tangan dibawah ini

Nama : Yustinus Andry Setiawan

NPM : 11 07 06659

Dengan ini menyatakan bahwa tugas akhir ini dengan judul "Pembangunan Perangkat Lunak Web Scraping untuk Situs Berita dan Peringkat Berita" merupakan hasil penelitian saya pada tahun akademik 2015/2016 yang bersifat originalitas dan tidak mengandung plagiasi dari karya manapun.

Bila dikemudian hari ditemukan ketidaksesuaian dengan pernyataan ini maka saya bersedia dituntut dan diproses dengan ketentuan yang berlaku termasuk dicabut gelar sarjana yang diberikan Universitas Atma Jaya Yogyakarta.

Demikian pernyataan ini dibuat dengan sesungguhnya dan dengan sebenar-benarnya.

Yogyakarta, 21 Oktober 2015

Yang menyatakan



Yustinus Andry Setiawan

HALAMAN PERSEMBAHAN

Karya Tugas Akhir ini kupersembahkan untuk :

Tuhan Yesus Kristus dan Bunda Maria

sebagai ungkapan rasa syukur dan pujianku,

untuk **Bapak** dan **Ibu** tersayang,

sebagai ungkapan kasih, sayang dan baktiku,

teman-teman semua yang terkasih,

TERIMA KASIH

"Barangsiapa ingin mutiara, harus berani terjun di lautan yang dalam."

Ir. Soekarno

KATA PENGANTAR

Puji syukur kepada Tuhan atas semua karunia dan berkat-Nya yang telah dilimpahkan kepada penulis sehingga penulis dapat menyelesaikan laporan skripsi ini dengan baik. Skripsi adalah studi akhir yang merupakan salah satu tugas akhir yang diwajibkan pada mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta setelah lulus mata kuliah teori, praktikum, dan kerja praktek. Tujuan dari pembuatan skripsi ini adalah sebagai salah satu syarat untuk mencapai derajat sarjana Teknik Informatika dari Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan skripsi ini tidak terlepas dari bantuan berbagai pihak yang telah menyumbangkan pikiran, tenaga, dukungan, bimbingan, dan doa kepada penulis baik secara langsung maupun tidak langsung. Oleh sebab itu, penulis mengucapkan terimakasih kepada :

1. Tuhan Yesus Kristus yang telah memberikan petunjuk dan harapan, serta melimpahkan karunia dan berkat-Nya kepada penulis.
2. Bapak Dr. A. Teguh Siswanto selaku Dekan Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
3. Bapak B. Yudi Dwiandiyanta, S.T., M.T., selaku Ketua Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.

4. Bapak Y. Sigit Purnomo W.P., S.T., M.Kom., selaku Dosen Pembimbing I yang telah menawarkan judul skripsi kepada penulis serta atas bimbingan, petunjuk, dan pengarahan hingga skripsi dapat diselesaikan dengan baik.
5. Bapak Kusworo Anindito, S.T., M.T., selaku Dosen Pembimbing II yang telah memberikan bimbingan, petunjuk, dan pengarahan kepada penulis sehingga skripsi dapat diselesaikan dengan baik.
6. Seluruh Staf Pengajar dan Laboran Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama masa kuliah di Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
7. Bapak dan ibu yang selalu mendoakan serta memberikan dorongan dan motivasi kepada penulis agar dapat menyelesaikan kuliah dengan baik dan tepat waktu.
8. Seluruh keluarga besar tercinta yang selalu memberikan dorongan dan semangat untuk penulis selama masa kuliah.
9. Teman dan sahabat Teknik Informatika angkatan 2011 Universitas Atma Jaya Yogyakarta, Firman, Vincent, Bangkit, Dimas, Theofilus, Xenix, Yanuar dan semuanya yang tidak bisa disebutkan satu per satu. Terimakasih atas bantuan, kebersamaan, kerja sama, canda tawa, suka duka dan senyuman kalian tidak dapat terlupakan.

10. Teman-teman KSL (Kelompok Studi Linux) yang telah memberikan pengalaman berorganisasi dan ilmunya tentang linux.

11. Teman-teman KKN 66 Padukuhan Palgading : Johan, Paul, Sahab, Tata, Oyen, Debora, Ayu yang telah membuat cerita yang tidak akan terlupakan selama 1 bulan. Terima kasih dukungan dan semangatnya.

12. Teman-teman dan pihak lain yang tidak mungkin disebutkan satu per satu yang telah membantu selama pengerjaan skripsi.

13. Semua orang yang secara tidak langsung memberikan dukungan dan semangat.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna karena keterbatasan waktu dan pengetahuan yang dimiliki penulis. Oleh karena itu segala kritik dan saran yang bersifat membangun sangat diharapkan. Akhir kata, semoga skripsi ini dapat berguna dan bermanfaat bagi semua pihak.

Yogyakarta, Oktober 2015

Penulis

**PEMBANGUNAN PERANGKAT LUNAK WEB SCRAPING UNTUK SITUS
BERITA DAN PERINGKAS BERITA**

Disusun oleh :

Yustinus Andry Setiawan

110706659

INTISARI

Informasi dengan mudah didapat melalui online atau internet. Salah satunya adalah informasi berupa berita. Situs Berita di Indonesia ini ada bermacam-macam. Banyaknya penyedia layanan berita tersebut membuat pembaca berita harus membuka beberapa situs untuk mencari berita yang berbobot. Selain harus membuka banyak situs berita, iklan-iklan dalam situs berita juga mengganggu. Pembaca menjadi tidak fokus membaca berita karena adanya iklan dalam situs berita. Biasanya pembaca harus menutup terlebih dahulu iklan dalam bentuk pop up pada situs berita. Hal tersebut membuat tersitanya waktu dan merepotkan untuk membaca berita.

Dengan permasalahan tersebut, penulis membuat aplikasi yang dapat mengambil data berita dari berbagai situs dan menampungnya di satu tempat. Aplikasi tersebut memanfaatkan teknik web scraping, yaitu teknik untuk mengambil informasi dari suatu website. Berita yang sudah didapatkan akan ditampilkan pada aplikasi mobile pembaca berita. Selain menampilkan data berita dari berbagai sumber, aplikasi mobile ini juga dapat meringkas berita. Metode yang digunakan untuk meringkas pada aplikasi ini adalah metode TF-IDF.

Dengan adanya aplikasi ini diharapkan memudahkan pembaca berita dalam membaca berita. Pembaca berita dapat menggunakan aplikasi mobile ini untuk membaca berita dari berbagai sumber dan dapat melihat rangkuman dari beberapa berita berdasarkan berita yang ingin dicari. Berita dapat dibaca dimana saja dan kapan saja dengan aplikasi mobile pembaca dan peringkas berita.

Kata Kunci : Web scraping, situs berita, aplikasi mobile, Android, peringkas berita

Pembimbing I : Y. Sigit Purnomo W.P., S.T., M.Kom.

Pembimbing II : Kusworo Anindito, S.T., M.T.

Jadwal Pendadaran : 21 Oktober 2015

DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR	iv
INTISARI	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Metodologi Penelitian.....	5
1.6 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA	8
BAB III LANDASAN TEORI	14
3.1 <i>Web server</i>	14
3.2 <i>Situs Berita Online</i>	15
3.3 <i>Web Service</i>	16
3.4 <i>JSON (Java Script Object Notation)</i>	18
3.5 <i>Web scraping</i>	18
3.6 <i>Automatic Text Summarization</i>	20
3.7 <i>Metode TF-IDF</i>	20
3.8 <i>Cron Job</i>	22
3.9 <i>Aplikasi Mobile</i>	23
BAB IV ANALISIS DAN PERANCANGAN PERANGKAT LUNAK	24

4.1	Analisis Latar Belakang Sistem.....	24
4.2	Analisis Sistem yang Akan Dibangun.....	25
4.2.1	Fungsi Produk	27
4.2.2	Use Case Diagram	29
4.2.3	Entity Relationship Diagram	30
4.3	Perancangan Sistem.....	30
4.3.1	Perancangan Arsitektur	31
4.3.2	Physical Data Model	33
4.3.3	Deskripsi Perancangan Antarmuka	34
4.3.3.1	Sketsa UI Aplikasi KUBERIN <i>Web</i> dan Deskripsinya	34
4.3.3.1.1	Antarmuka Halaman Login <i>Web</i>	34
4.3.3.1.2	Antarmuka Halaman <i>Scraping</i>	35
4.3.3.1.3	Antarmuka Halaman Lihat Data Berita.....	36
4.3.3.1.4	Antarmuka Halaman <i>Update Data Tags</i>	37
4.3.3.1.5	Antarmuka Halaman Pengelolaan Kategori...	38
4.3.3.1.6	Antarmuka Halaman <i>Summarize Berita Web</i> ...	40
4.3.3.2 Sketsa UI Aplikasi KUBERIN <i>Mobile</i> dan Deskripsinya	41
4.3.3.2.1	Antarmuka Halaman Kategori Berita.....	41
4.3.3.2.2	Antarmuka Halaman <i>Summarize Berita Mobile</i>	42
4.3.3.2.3	Antarmuka Halaman <i>Trending Topik</i>	43
BAB V	IMPLEMENTASI DAN PENGUJIAN SISTEM	44
5.1	Definisi Sistem.....	44
5.2	Implementasi Sistem.....	52
5.2.1	Antarmuka Aplikasi <i>Web</i> Kuberin	52
5.2.1.1	Antarmuka Halaman Login Web	52
5.2.1.2	Antarmuka Halaman <i>Scraping</i>	54
5.2.1.3	Antarmuka Halaman Lihat Data Berita	66
5.2.1.4	Antarmuka Halaman <i>Update Data Tags</i>	68
5.2.1.5	Antarmuka Halaman Pengelolaan Kategori	71

5.2.1.6 Antarmuka Halaman Summarize Berita	75
5.2.2 Antarmuka Aplikasi <i>Mobile</i> Kuberin	83
5.2.2.1 Antarmuka Halaman Kategori Berita	83
5.2.2.2 Antarmuka Halaman <i>Summarize</i> Berita <i>Mobile</i> ...	92
5.2.2.3 Antarmuka Halaman Trending Topik	101
5.2.3 Aplikasi Web Scraper	104
5.2.3.1 Mengambil URL Berita	104
5.2.3.2 Mengambil Data Berita	106
5.2.3.3 Mengecek <i>Tags</i> Tidak Aktif	109
5.3 Hasil Pengujian Perangkat Lunak.....	112
5.4 Hasil Pengujian Terhadap Pengguna.....	123
5.4.1 Pengujian Fungsionalitas Aplikasi KUBERIN Mobile	123
5.4.2 Pengujian Antarmuka Aplikasi KUBERIN Mobile ...	131
5.5 Analisis Kelebihan dan Kekurangan Aplikasi.....	139
BAB VI KESIMPULAN DAN SARAN	141
6.1 Kesimpulan.....	141
6.2 Saran.....	142
DAFTAR PUSTAKA	143
LAMPIRAN	146

DAFTAR GAMBAR

Gambar 3.1	Arsitektur <i>Web server</i>	14
Gambar 3.2	Contoh Situs Berita <i>Online</i>	15
Gambar 3.3	Penggunaan JSON REST	18
Gambar 3.4	Proses <i>Web Scraping</i>	19
Gambar 3.5	Contoh Operasi <i>Cron Job</i>	22
Gambar 4.1	Arsitektur Perangkat lunak KUBERIN.....	26
Gambar 4.2	Use Case Diagram KUBERIN	29
Gambar 4.3	Entity Relationship Diagram KUBERIN	30
Gambar 4.4	Perancangan Arsitektur KUBERIN <i>Web Application</i>	31
Gambar 4.5	Perancangan Arsitektur KUBERIN <i>Mobile Application</i>	32
Gambar 4.6	Physical Data Model KUBERIN	33
Gambar 4.7	Antarmuka Login <i>Web</i>	34
Gambar 4.8	Antarmuka <i>Scraping</i>	35
Gambar 4.9	Antarmuka Lihat Data Berita	36
Gambar 4.10	Antarmuka <i>Update Data Tags</i>	37
Gambar 4.11	Antarmuka Pengelolaan Kategori	38
Gambar 4.12	Antarmuka <i>Summarize Berita Web</i>	40
Gambar 4.13	Antarmuka Kategori Berita	41
Gambar 4.14	Antarmuka <i>Summarize Berita Mobile</i>	42
Gambar 4.15	Antarmuka <i>Trending Topik</i>	43
Gambar 5.1	Antarmuka Halaman Login <i>Web</i>	52
Gambar 5.2	Proses Menampilkan Halaman Login.....	52
Gambar 5.3	Proses Login <i>Web</i>	53
Gambar 5.4	Antarmuka Halaman <i>Scraping</i>	54
Gambar 5.5	Proses <i>Test Scraping</i>	55
Gambar 5.6	Cuplikan code untuk mendapatkan atau membuka halaman <i>web</i>	56

Gambar 5.7 Cuplikan code untuk mengambil URL berita	56
Gambar 5.8 Cuplikan code untuk mengambil judul berita.....	58
Gambar 5.9 Cuplikan code untuk mengambil kategori berita.....	59
Gambar 5.10 Cuplikan code untuk mengambil tanggal dan waktu berita.....	60
Gambar 5.11 Cuplikan code untuk mengambil URL Gambar.....	61
Gambar 5.12 Cuplikan code untuk mengambil isiberita	61
Gambar 5.13 Cuplikan code untuk membuang konten tidak perlu	62
Gambar 5.14 JSON <i>Output Fungsi Test</i>	65
Gambar 5.15 Proses Simpan <i>Scraping</i>	66
Gambar 5.16 Antarmuka Halaman Lihat Data Berita.....	66
Gambar 5.17 Proses Menampilkan halaman lihat data berita.....	67
Gambar 5.18 Proses Menampilkan detail berita.....	68
Gambar 5.19 Antarmuka Halaman <i>Update Data Tags</i>	68
Gambar 5.20 Proses Menampilkan Halaman <i>Update Data Tags</i>	69
Gambar 5.21 Proses Mengambil <i>Data Tags</i>	70
Gambar 5.22 Proses <i>Update Data Tags</i>	71
Gambar 5.23 Antarmuka Halaman Pengelolaan Kategori.....	71
Gambar 5.24 Proses Menampilkan halaman pengelolaan kategori	72
Gambar 5.25 Antarmuka Halaman <i>Summarize Berita Versi 1</i>	75
Gambar 5.26 Proses <i>Summarize Berita</i>	76
Gambar 5.27 Cuplikan Code Memasukkan Isi Berita Ke dalam <i>array</i>	77
Gambar 5.28 Cuplikan Code Pemanggilan <i>Single Summary</i>	77
Gambar 5.29 Cuplikan Code Pemanggilan <i>Multi Summary</i> , skor kalimat, dan hasil <i>summary</i>	78
Gambar 5.30 Cuplikan Code untuk mendapatkan data berita terkait.....	80
Gambar 5.31 Antarmuka Halaman <i>Summarize Berita Versi 2</i>	82

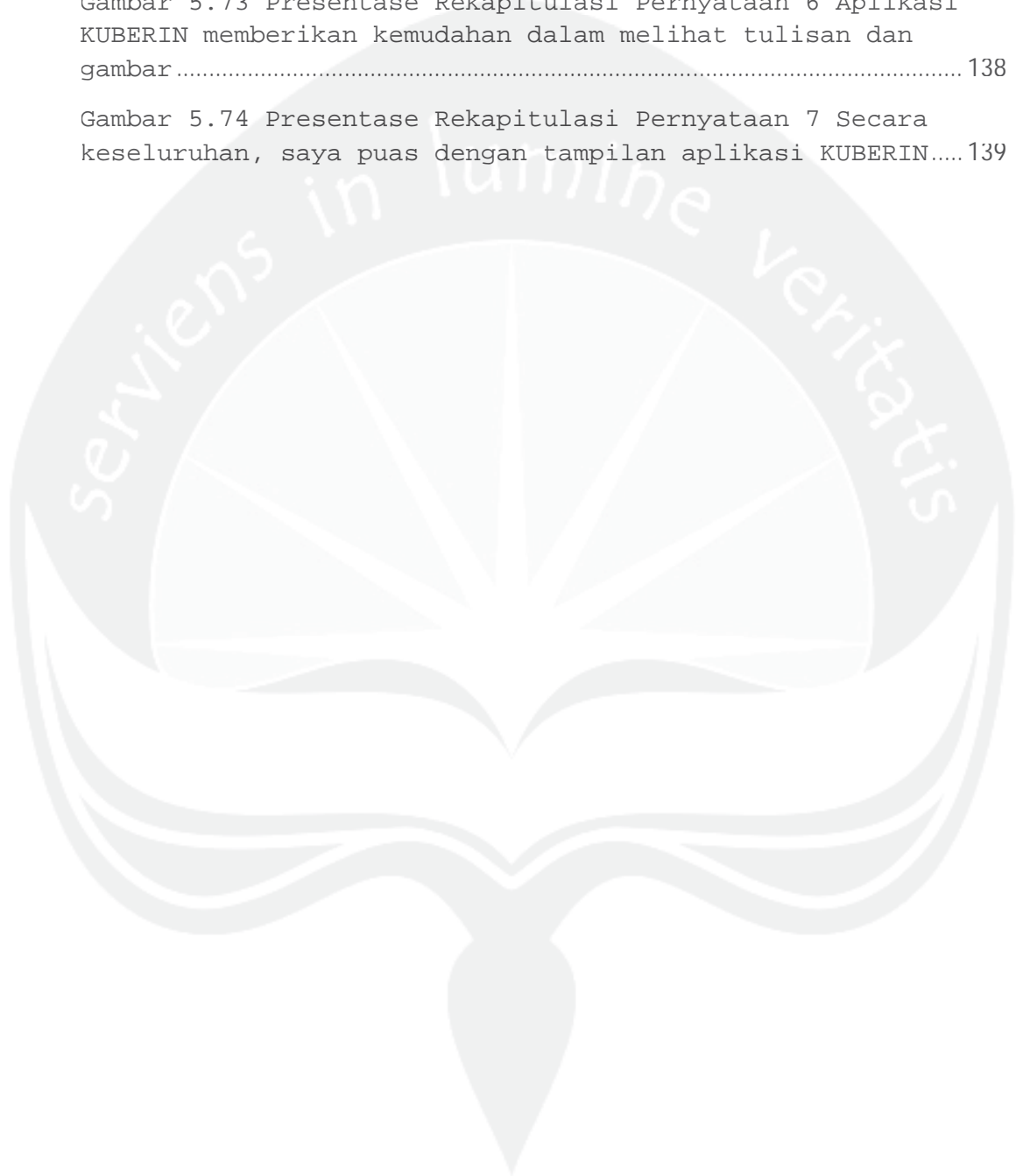
Gambar 5.32 Antarmuka Halaman Awal Kategori Berita.....	83
Gambar 5.33 Antarmuka Halaman <i>List</i> Berita	84
Gambar 5.34 Proses Mendapatkan data berita.....	85
Gambar 5.35 Fungsi kategori berita.....	86
Gambar 5.36 Fungsi GetJumlahDataBeritaKategori dan fetchBeritaKategori.....	87
Gambar 5.37 JSON <i>Output</i> fungsi kategori.....	88
Gambar 5.38 Fungsi <code>doInBackground</code> pada class <code>AsyncTaskNewsParseJson</code>	88
Gambar 5.39 Fungsi <code>onPostExecute</code> pada class <code>AsyncTaskNewsParseJson</code>	89
Gambar 5.40 Cuplikan Code <code>itemClick</code> Detail Berita.....	90
Gambar 5.41 Antarmuka Halaman Detail Berita dan <i>Web View</i> Berita.....	91
Gambar 5.42 Antarmuka Halaman <i>Summarize</i> Berita	92
Gambar 5.43 Proses Mendapatkan Data Sumber berita.....	93
Gambar 5.44 Proses <i>Summarize</i> Berita <i>Button Summarize</i>	94
Gambar 5.45 JSON <i>Output</i> fungsi <code>summarizeberita</code>	95
Gambar 5.46 Fungsi <code>doInBackground</code> pada class <code>AsyncTaskSummarize</code>	96
Gambar 5.47 Antarmuka Halaman <i>Tab Summary</i> , <i>Tab</i> Berita Terkait, dan <i>Tab</i> Skor.....	97
Gambar 5.48 Proses <i>Summarize</i> Berita <i>Button</i> Filter Berita.....	98
Gambar 5.49 JSON <i>Output</i> fungsi <code>getDataBeritaSummarize</code>	99
Gambar 5.50 Antarmuka Halaman Pilih Berita.....	100
Gambar 5.51 Antarmuka Halaman <i>Trending</i> Topik.....	101
Gambar 5.52 Proses Mendapatkan Data <i>Trending</i> Topik.....	102
Gambar 5.53 Proses Mendapatkan Data Detail <i>Trending</i> Topik	103
Gambar 5.54 Cuplikan code Fungsi <code>doInBackground</code> <code>AsyntaskGetHasilTrending</code>	103
Gambar 5.55 Proses <i>Cron</i> URL Berita.....	104

Gambar 5.56 <i>Command</i> untuk menjalankan script cronurlberita	105
Gambar 5.57 Proses Cron Berita.....	107
Gambar 5.58 <i>Command</i> untuk menjalankan script cronberita.....	107
Gambar 5.59 Proses <i>Cron Cek Tags</i> Tidak Aktif.....	109
Gambar 5.60 <i>Command</i> untuk menjalankan script croncekaktif	110
Gambar 5.61 Presentase Rekapitulasi Pernyataan 1 Aplikasi KUBERIN memberikan kemudahan dalam membaca berita/melihat sebuah berita dari kumpulan berita berbagai sumber.....	125
Gambar 5.62 Presentase Rekapitulasi Pernyataan 2 Data berita yang ditampilkan pada aplikasi KUBERIN isinya lengkap.....	126
Gambar 5.63 Presentase Rekapitulasi Pernyataan 3 Data berita yang ditampilkan pada aplikasi KUBERIN memiliki kesesuaian/saling berkaitan antara judul, gambar, dan berita	127
Gambar 5.64 Presentase Rekapitulasi Pernyataan 4 Aplikasi KUBERIN memberikan kemudahan dalam mencari berita berdasarkan kategori yang tersedia.....	128
Gambar 5.65 Presentase Rekapitulasi Pernyataan 5 Aplikasi KUBERIN memberikan kemudahan untuk mendapatkan ringkasan berita berdasarkan topik yang diinginkan pengguna.....	129
Gambar 5.66 Presentase Rekapitulasi Pernyataan 6 Aplikasi KUBERIN fitur summarize memberikan ringkasan berita yang kualitasnya sudah baik.....	130
Gambar 5.67 Presentase Rekapitulasi Pernyataan 7 Secara keseluruhan, saya puas dengan kinerja aplikasi KUBERIN.....	131
Gambar 5.68 Presentase Rekapitulasi Pernyataan 1 Aplikasi KUBERIN memiliki tampilan yang menarik.....	133
Gambar 5.69 Presentase Rekapitulasi Pernyataan 2 Aplikasi KUBERIN mudah dipahami fungsionalitasnya dari antarmuka aplikasi.....	134
Gambar 5.70 Presentase Rekapitulasi Pernyataan 3 Aplikasi KUBERIN mudah untuk digunakan.....	135
Gambar 5.71 Presentase Rekapitulasi Pernyataan 4 Cara penggunaan aplikasi KUBERIN ini mudah dipelajari.....	136

Gambar 5.72 Presentase Rekapitulasi Pernyataan 5 Desain warna aplikasi KUBERIN nyaman untuk dilihat dan komposisi warnanya pas..... 137

Gambar 5.73 Presentase Rekapitulasi Pernyataan 6 Aplikasi KUBERIN memberikan kemudahan dalam melihat tulisan dan gambar..... 138

Gambar 5.74 Presentase Rekapitulasi Pernyataan 7 Secara keseluruhan, saya puas dengan tampilan aplikasi KUBERIN..... 139



DAFTAR TABEL

Tabel 2.1 Tabel Perbandingan Sistem	12
Tabel 5.1 Tabel File KUBERIN <i>Web</i>	45
Tabel 5.2 Tabel File KUBERIN <i>Mobile</i>	48
Tabel 5.3 Hasil Pengujian Perangkat Lunak	112
Tabel 5.4 Tabel Pengujian Fungsionalitas Apikasi KUBERIN Mobile	123
Tabel 5.5 Tabel Pengujian Antarmuka Apikasi KUBERIN Mobile	131

DAFTAR LAMPIRAN

- I. Spesifikasi Kebutuhan Perangkat Lunak Kumpulan Berita Indonesia (SKPL KUBERIN)
- II. Deskripsi Perancangan Perangkat Lunak Kumpulan Berita Indonesia (DPPL KUBERIN)

