

PROGRAM STUDI TEKNIK INFORMATIKA

**PENERAPAN *WEB SCRAPING* UNTUK MENDAPATKAN BERITA
*ONLINE***

**SODRI SIBARANI
15142049P**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana Komputer
di Universitas Bina Darma**



**FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA DARMA
2019**



**PENERAPAN *WEB SCRAPING* UNTUK MENDAPATKAN BERITA
*ONLINE***

**SODRI SIBARANI
15142049P**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana Komputer
Di Universitas Bina Darma**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA DARMA
2019**

HALAMAN PENGESAHAN

**PENERAPAN WEB SCRAPING UNTUK MENDAPATKAN
BERITA ONLINE**

**SODRI SIBARANI
15142049P**


**Telah diterima salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Teknik Informatika**

Pembimbing I



Megawaty, M. Kom

Pembimbing II



Hadi Syaputra, M.Kom

Palembang, Juli 2019
Fakultas Ilmu Komputer
Universitas Bina Darma
Dekan,



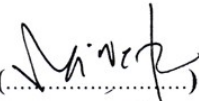



Dedy Syamsuar, S.Kom., M.I.T., Ph.D

HALAMAN PERSETUJUAN

Skripsi berjudul “PENERAPAN WEB SCRAPING UNTUK MENDAPATKAN BERITA ONLINE” Oleh “Sodri Sibarani”, telah dipertahankan di depan komisi penguji pada hari **Senin** Tanggal **29 Juli 2019**.

Komisi Penguji

1. **Ketua** **Megawaty, M.Kom.** 
(.....)
2. **Anggota** **Hadi Syaputra, M.Kom** 
(.....)
3. **Anggota** **Susan Dian Purnamasari, M.Kom** 
(.....)
4. **Anggota** **Andri, M.Cs** 
(.....)

Mengetahui,
Program Studi Teknik Informatika
Fakultas Ilmu Komputer
Universitas Bina Darma
Ketua,


Universitas **Bina Darma** 
Fakultas Ilmu Komunikasi

A. Haidar Mirza, S.T., M.Kom

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : SODRI SIBARANI

NIM : 15142049P

Dengan ini menyatakan bahwa :

1. Karya tulis saya (tugas akhir/skripsi/tesis) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (ahli madya/sarjana/magister) di Universitas Bina Darma atau diperguruan tinggi lain;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya sendiri dengan arahan tim pembimbing;
3. Didalam karya tulis ini karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukan ke dalam daftar rujukan;
4. Saya bersedia tugas akhir/skripsi/tesis, yang saya hasilkan dicek keasliannya menggunakan *plagirism checker* serta di unggah ke internet sehingga dapat diakses oleh publik secara daring;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan peraturan dan perundang-undangan.

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 29 Juli 2019

METERAI
PENCAP
TGL 20
FGA1AFF91370827
6000
ENAM RIBU RUPIAH
SODRI SIBARANI



NIM : 15142049P

MOTTO DAN PERSEMBAHAN

MOTTO

“Jadilah manusia berguna di manapun berada dan berusaha tidak akan menyombongkan diri dalam kegunaannya”

PERSEMBAHAN

Kupersembahkan Kepada :

- *Ayah dan Ibu Tercinta yang selalu mendoakan dan mengorbankan segalanya untuk keberhasilan ku.*
- *Para pendidikku*
- *Untuk sahabatku yang telah memberikan semangat dan membantu*
- *Kepada Dosen pembimbing skripsi ini.*
- *Almamater*

ABSTRAK

Web Scraping merupakan teknik untuk mengambil informasi dalam suatu situs *website* secara otomatis. Fokus dari aplikasi *Web Scraping* adalah mengambil informasi dan mengekstrak informasi, Pengindeksan *website* mempunyai hubungan dengan *WebScraping*, tetapi *Web Scraping* fokus pada transformasi website tidak terstruktur menjadi format data terstruktur, Format data terstruktur tersebut dapat disimpan dan dapat dianalisa di database, Berita dari berbagai situs dapat diambil informasinya dengan teknik *Web Scraping* dan disimpan dalam *database*, Selain aplikasi pengambil informasi dari berbagai situs, dibutuhkan juga *summarize* berita untuk mendapatkan hasil *summary* beberapa berita menjadi satu.

Permasalahan bahwa semua berita tidak terangkum dengan sesuai kriteria nya dan menyulitkan pembaca berita dalam mencari berita online sehingga orang sangat susah dalam mencari berita sesuai kriteria masing-masing, maka dari itu membutuhkan aplikasi untuk dalam memilah berita di setiap situs yang dikunjungi.

Dari penjelasan di atas maka penelitian ini mencari situs berita untuk mengumpulkan berita dari berbagai situs dan meringkas berita. Maka dalam penulisan skripsi ini diangkatlah sebuah judul yaitu "*Penerapan Web Scraping Berita Online*".

Kata Kunci : *web scraping, Berita Online*

KATA PENGANTAR



Puji dan syukur penulis panjatkan kepada Allah SWT karena berkat rahmat-Nya jualah peneliti dapat menyelesaikan skripsi ini dengan judul **“PENERAPAN WEB SCRAPING UNTUK MENDAPATKAN BERITA ONLINE”**.

Pada kesempatan yang baik ini, tak lupa penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat dan pemikiran dalam penyusunan skripsi ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd., M.M, selaku Rektor Universitas Bina Darma Palembang.
2. Dedy Syamsuar, S.Kom., M.I.T., Ph.D, selaku Dekan Fakultas Ilmu Komputer.
3. A. Haidar Mirza, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika.
4. Megawaty, M.Kom., selaku dosen pembimbing Iyang telah memberikan bimbingan dalam penelitian ini.
5. Hadi Syaputra, M.Kom., selaku dosen pembimbing Iyang telah memberikan bimbingan dalam penelitian ini.
6. Orang tua, serta keluargaku tercinta terima kasih atas do'a, restu, dan semangat sehingga dapat menyelesaikan proposal penelitian ini.

Dalam penulisan skripsi ini, penulis menyadari bahwa masih banyak kekurangan serta masih jauh dari sempurna. Maka dari itu peneliti mengharapkan masukan saran dan kritik yang membangun, untuk melengkapi kekurangan yang ada.

Palembang, Agustus 2019
Penulis

SodriSibarani

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PERNYATAAN.....	iv
PERSEMBAHAN DAN MOTTO.....	v
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian.....	3
1.4.1 Tujuan Penelitian.....	3
1.4.2 Manfaat Penelitian.....	3
1.5 Alat dan Bahan.....	4
1.6 Metode Pengumpulan Data.....	5
1.7 Proses <i>Web Scraping</i>	5
BAB II TINJAUAN PUSTAKA.....	8
2.1. Landasan Teori.....	8
2.1.1 Search Engine Result Pages	8
2.1.2 MySql.....	9
2.1.3 Search Engine Optimization.....	9

2.1.4 <i>Web Scraping</i>	10
2.1.4.1 Domain	11
2.1.4.2 Domain Authority.....	12
2.1.4.3 Regular Expression.....	13
2.2 Penelitian Sebelumnya.....	17
BAB III ANALISIS DAN PERANCANGAN.....	19
3.1 Observasi.....	19
3.2 Pentaho.....	19
3.3 Pengambilan Gambar Dari Halaman Detail.....	20
3.4 Rancangan Basis Data.....	21
3.5 Halaman Java Script.....	22
3.6 AlurSistem Web Scraping.....	23
3.7 Pentaho Report Designer.....	24
BAB IV HASIL DAN PEMBAHASAN.....	26
4.1. Hasil.....	26
4.2. Pembahasan.....	26
4.2.1. Halaman web liputan6.com.....	26
4.2.2. Update Berita.....	27
4.2.3. HalamanAplikasi.....	29
BAB V KESIMPULAN DAN SARAN.....	39
5.1. Kesimpulan.....	39
5.2. Saran.....	40
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

1.1	Ilustrasi Cara Kerja Scraping.....	17
3.1	Aliran Proses Web Scraping	22
3.2	prompt Service	24
4.1	Web Liputan6.com.....	27
4.2	Update Berita.....	28
4.3	Halaman Menu Aplikasi.....	28
4.4	Gambar.....	29
4.5	Gambar.....	29
4.6	Gambar.....	30
4.7	Gambar.....	30
4.8	Gambar.....	31
4.9	Gambar.....	31
4.10	Gambar.....	32
4.11	Gambar.....	32
4.12	Gambar.....	33
4.13	Gambar.....	33
4.13	Gambar.....	34
4.14	Gambar.....	34
4.15	Gambar.....	35
4.16	Gambar.....	35
4.17	Hasil Scraping.....	36
4.18	Hasil Scraping.....	36
4.19	Hasil Scraping.....	37
4.20	Hasil Scraping.....	38

DAFTAR TABEL

Tabel	Halaman
3.1 list Kode Regex.....	14