

SKRIPSI

**TEKNIK WEB SCRAPING UNTUK PENCARIAN PRODUK
PROMO TOKO ONLINE**

***WEB SCRAPING TECHNIQUE FOR ONLINE SHOP PROMO
PRODUCT SEARCH***



OBET EDM MARTINUS FRANCISCO

135610087

**PROGRAM STUDI SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN
KOMPUTER AKAKOM
YOGYAKARTA**

2019

SKRIPSI

**TEKNIK WEB SCRAPING UNTUK PENCARIAN PRODUK
PROMO TOKO ONLINE**

***WEB SCRAPING TECHNIQUE FOR ONLINE SHOP PROMO
PRODUCT SEARCH***

Diajukan sebagai salah satu syarat untuk menyelesaikan studi jenjang strata

satu (S1)

Program Studi Sistem Informasi

Sekolah Tinggi Manajemen Informatika dan Komputer

AKAKOM

Yogyakarta

Disusun Oleh :

OBET EDOM MARTINUS FRANCISCO

135610087

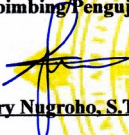
**PROGRAM STUDI SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMAN INFORMATIKA DAN
KOMPUTER AKAKOM
YOGYAKARTA**

2019

HALAMAN PENGESAHAN
SKRIPSI
TEKNIK WEB SCRAPING UNTUK PENCARIAN PRODUK
PROMO TOKO ONLINE

Telah dipersiapkan dan disusun oleh
OBET EDOM MARTINUS FRANCISCO
135610087

Telah dipertahankan didepan Tim Penguji
Pada tanggal 28 januari 2019
Susunan Tim Penguji

Pembimbing/Penguji  <u>Fx. Henry Nugroho, S.T., M.Cs.</u> NIP/NPP. 19800121 2005 01 1001	Ketua Penguji  <u>Pulut Suryati, S.Kom., M. Cs.</u> NIP/NPP. 19780315 2005 01 2002
	Anggota  <u>Emy Susanti, S. Kom., M. Cs.</u> NIP/NPP. 19790303 2005 01 2001

Skripsi ini telah diterima sebagai salah satu persyaratan untuk
memperoleh gelar Sarjana Komputer

Tanggal **11 FEB 2019**

Ketua Program Studi Sistem Informasi
Deborah Kurniawati, S.Kom., M.Cs.
NIP/NPP. 051149

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa Laporan Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Ahli Madya/kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, Januari 2019



(OBET EDM MARTINUS FRANCISCO)

PERSEMBAHAN

Skripsi ini, saya persembahkan untuk :

- Tuhan Yesus Kristus yang selalu memberikan rahmat kepada saya sehingga dapat menyelesaikan skripsi ini.
- Bapak dan Ibu saya yang selalu memberikan dukungan yang tidak terbatas.
- Semua saudara saya yang telah memberikan motivasi yang begitu bermanfaat.
- Terima kasih kepada Dosen pembimbing Fx. Henry Nugroho, S.T., M. Cs. yang telah membimbing saya dan membantu terselesaikannya skripsi ini.
- Teman-teman yang selalu memberikan semangat dan doa kepada saya.
- Semua orang yang tidak tersebut namanya di halaman ini yang membantu saya dalam menyelesaikan tugas skripsi ini.

MOTTO

“Mintalah, maka akan diberikan kepada-Mu. Carilah, maka kamu akan mendapat. Ketuklah maka pintu akan dibukakan bagi-Mu”

“Jika Tuhan di pihak kita semua ada jalan keluarnya”

“Kesederhanaan itu mengalahkan segalanya”

“Santai tapi hasilnya itu pasti”

“Jangan tunda sampai besok apa yang bisa engkau kerjakan hari ini”

“Hari ini harus lebih baik dari hari kemarin dan hari esok adalah harapan”

“Belajar akan memberikan kita pengetahuan dan pemahaman baru, sehingga kita mampu menghadapi tantangan baru yang membentang di depan kita”

DAFTAR ISI

COVER	
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN.....	iii
PERSEMBAHAN.....	iv
MOTTO.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	x
KATA PENGANTAR.....	xi
ABSTRAK.....	xii
<i>ABSTRACT</i>	xiii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Ruang Lingkup.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB 2 TINJAUAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Tinjauan Pustaka.....	6

2.2 Dasar Teori.....	9
2.2.1 Web Scraping.....	9
2.2.2 <i>Codeigniter</i>	9
2.2.3 <i>Parsing</i>	10
2.2.4 <i>PHP</i>	10
2.2.5 <i>CURL</i>	11
2.2.6 <i>Simple Html DOM</i>	11
BAB 3 METODE PENELITIAN.....	13
3.1 Deskripsi Sistem.....	13
3.2 Tahapan Pembangunan Sistem.....	14
3.2.1 Mencari Referensi.....	14
3.2.2 Perancangan Sistem.....	14
3.2.3 Pengguna Sistem.....	14
3.2.4 Kebutuhan Data.....	15
3.2.5 Kebutuhan Proses.....	15
3.2.6 Arsitektur Sistem.....	15
3.2.7 <i>Flowchart</i> Alur Sistem.....	16
3.3 Pembuatan Sistem.....	18
3.4 Pengujian.....	18
3.5 Rancangan Masukan dan Keluaran.....	18
BAB 4 IMPLEMENTASI DAN PEMBAHASAN SISTEM.....	21
4.1 Implementasi	21
4.1.1 Konfigurasi <i>Database</i>	21

4.1.2 Menambahkan Url Promo.....	22
4.1.3 Otentifikasi <i>User</i>	23
4.1.4 <i>Scraping</i>	24
4.2 Pembahasan	25
4.2.1 Pengambilan Tag Html Dari Web Target.....	26
4.2.2 Website Tidak Berhasil.....	26
4.2.3 Pengujian Sistem.....	28
BAB 5 PENUTUP.....	35
5.1 Kesimpulan.....	35
5.2 Saran.....	35
DAFTAR PUSTAKA.....	37
LAMPIRAN	

DAFTAR GAMBAR

Gambar 3.1 Arsitektur Sistem.....	15
Gambar 3.2 <i>Flowchat</i> Alur Sistem.....	16
Gambar 3.3 Rancangan Halaman Login.....	19
Gambar 3.4 Halaman Output Admin.....	19
Gambar 3.5 Halaman Pengguna.....	20
Gambar 4.1 Konfigurasi <i>Database</i>	21
Gambar 4.2 Menambahkan Url promo.....	22
Gambar 4.3 Otentifikasi <i>User</i>	23
Gambar 4.4 <i>Scraping</i>	24
Gambar 4.5 Pengambilan tag html.....	26
Gambar 4.6 <i>Url</i> Target Gagal	27
Gambar 4.7 <i>Url</i> Target Gagal pada website shopee.....	27
Gambar 4.8 Promo Tokopedia.....	28
Gambar 4.9 Promo Lazada.....	29
Gambar 4.10 Promo Bukalapak.....	30

DAFTAR TABEL

Tabel 2.1 Tabel Perbandingan Penelitian.....	8
Tabel 3.1 Nama Marketplace.....	13
Tabel 4.1 Marketplace Yang Berhasil.....	25
Tabel 4.2 Marketplace Yang Tidak Berhasil.....	25
Tabel 4.3 Perbandingan Website Tokopedia.....	31
Tabel 4.4 Perbandingan Website Lazada.....	32
Tabel 4.5 Perbandingan Website Bukalapak.....	33

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Kuasa karena banyaknya limpahan karunia dan nikmatnya sehingga penulis dapat menyelesaikan skripsi ini dengan lancar tanpa kendala yang berarti.

Dalam penyusunan naskah skripsi ini tidak akan terlaksana tanpa dukungan, bimbingan dan petunjuk dari semua pihak yang telah membantu sehingga naskah skripsi dapat terselesaikan dengan baik. Penulis mengucapkan terimakasih kepada:

1. Bapak Ir. Totok Suprawoto, M.M., M.T. selaku ketua Sekolah Tinggi Manajemen Informatika dan Komputer AKAKOM Yogyakarta.
2. Ibu Deborah Kurniawati, S.Kom., M.Cs., selaku Ketua Program Studi Sistem Informasi.
3. Bapak Fx. Henry Nugroho, S.T., M. Cs. selaku pembimbing skripsi yang telah banyak memberikan motivasi, bantuan dan pengarahan.
4. Seluruh dosen di Jurusan Sistem informasi STMIK AKAKOM YOGYAKARTA yang selama ini telah membagikan ilmunya.
5. Ibu dan bapak yang telah memberikan seluruh jiwa raganya dan waktunya untuk mendukung, mendoakan dan berkerja keras demi anaknya.
6. Teman-teman angkatan 2014 yang terlalu banyak jika di sebutkan satu persatu disini.

Dalam penyusunan naskah skripsi ini, penulis menyadari masih banyak kekurangan. Akhir kata semoga Karya Tulis ini dapat memberikan banyak manfaat bagi kita semua.

Yogyakarta

Penulis

ABSTRAK

Search engine yaitu kombinasi perangkat keras dan perangkat lunak komputer yang disediakan oleh perusahaan tertentu melalui website yang telah ditentukan. Search engine mengumpulkan informasi dari web melalui program web *crawler* yang secara periodik menelusuri web. Proses pengambilan informasi dari website-website yang ada ini disebut dengan “*web scraping*”. *Web Scraping* adalah suatu teknik penggalian informasi dari situs web.

Web Scraping berkaitan erat dengan pengindeksan web, adapun cara mengembangkan teknik web scraping yaitu dengan cara pembuat program mempelajari dokumen HTML dari website yang akan diambil informasinya untuk di tag HTML tujuannya ialah untuk mengapit informasi yang diambil setelah itu pembuat program mempelajari teknik *scraping* pada website yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraping* yang akan kita buat. Perlu pula diperhatikan bahwa implementasi scraping pada tulisan ini hanya melibatkan mesin pencari promo dari toko online seperti: Tokopedia, Lazada, dan Bukalapak.

Kata Kunci : *Web Scraping, Crawling, Promo, Simple Html DOM*

ABSTRACT

Search engines are a combination of hardware and computer software provided by a particular company through a predetermined website. Search engines collect information from the web through a web crawler program that periodically searches the web. The process of retrieving information from existing websites is called "web scraping." Web Scraping is a technique of extracting information from a website.

Web scraping is closely related to indexing the web, as for how to develop web scraping techniques by means of program makers learning HTML documents from websites that will be taken for HTML tags. information is taken to be imitated on the web scraping application that we will make. It should also be noted that the implementation of scraping in this paper only involves promo search engines from online stores such as: Tokopedia, Lazada, and Bukalapak.

Keywords :*Web Scraping, Crawling, Promos, Simple DOM XML*