

**IMPLEMENTASI *WEBSCRAPING* UNTUK PENCARIAN
PRODUK DENGAN METODE *EXTENSIBLE PATH
LANGUAGE (XPATH)* PADA HALAMAN WEBSITE
*E-COMMERCE***

SKRIPSI

**Untuk Memenuhi Persyaratan
Memperoleh Gelar Sarjana**

**Oleh :
Rifqi Rosyidi
NIM 16.51.0010**



PROGRAM STUDI S1 - SISTEM INFORMASI

**KEMENTERIAN RISET TEKNOLOGI DAN PENDIDIKAN TINGGI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
PRADNYA PARAMITA
MALANG
2020**

LEMBAR PERSETUJUAN

**IMPLEMENTASI *WEBCRAPING* UNTUK PENCARIAN
PRODUK DENGAN METODE *EXTENSIBLE PATH*
LANGUAGE (XPATH) PADA HALAMAN WEBSITE *E-*
*COMMERCE***

Oleh:
Rifqi Rosyidi
16.51.0010

Telah disetujui oleh pembimbing untuk diujikan
Pada tanggal 10 Juni 2020

Menyetujui,
Komisi Pembimbing

Pembimbing I



Dr. Tubagus Mohamad Akhriza, S.Si., MMSI., Ph.D.
NIK 00.12.01.002

Pembimbing II



Dwi Safiroh Utsalina, S.Kom., MMSI.
NIK. 00.12.01.016

Mengetahui,
Ketua Program Studi S1 - Sistem Informasi



Dwi Safiroh Utsalina, S.Kom., MMSI.
NIK. 00.12.01.016

LEMBAR PENGESAHAN

IMPLEMENTASI *WEBSCRAPING* UNTUK PENCARIAN
PRODUK DENGAN METODE *EXTENSIBLE PATH
LANGUAGE (XPATH)* PADA HALAMAN WEBSITE *E-
COMMERCE*

Skripsi oleh Rifqi Rosyidi
Telah dipertahankan di depan dewan penguji
Pada tanggal 18 Juni 2020

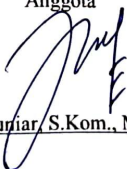
Dewan Penguji,
Ketua


Mahmud Yunus, S.Kom., M.Pd. MT.
NIK. 00.12.01.011

Anggota

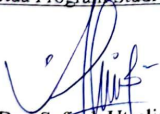

Dr. Tubagus Mohamad Akhriza, S.Si., MMSI., Ph.D.
NIK. 00.12.01.002

Anggota


Eka Yuniar, S.Kom., MMSI.

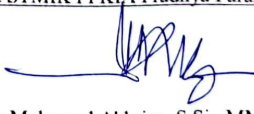
Mengetahui,

Ketua Program Studi SI - Sistem Informasi


Dwi Saffron Utsalina, S.Kom., MMSI.
NIK. 00.12.01.016

Mengesahkan,

Ketua STMIK PPKIA Pradnya Paramita


Dr. Tubagus Mohamad Akhriza, S.Si., MMSI., Ph.D.
NIK. 00.12.01.002

PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam Naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu Perguruan Tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No 20 Tahun 2003, Pasal 25 ayat 2 dan pasal 70).

Malang, 20 Mei 2020
Mahasiswa,

Materai 10k

Nama : Rifqi Rosyidi
NIM : 16.51.0010

ABSTRAK

Rosyidi, Rifqi. 2020. Implementasi *Webscraping* untuk Pencarian Produk Dengan Metode *Extensible Path Language* (XPath) Pada Halaman Website *E-Commerce*. Skripsi, Program Studi Sistem Informasi STMIK PPKIA Pradnya Paramita. Pembimbing: (I) Dr. Tubagus Mohamad Akhriza, S.Si., MMSI., (II) Dwi Safiroh Utsalina, S.Kom., MMSI.

Kata Kunci: *Webscraping*, *Webdata Extraction*, *XPath*, *E-Commerce*.

Information Overload yaitu banyaknya jumlah informasi yang diterima oleh manusia sehingga menimbulkan kesulitan dalam penerimaan dan pengolahan informasi. Fenomena ini juga terjadi salah satunya pada situs jual beli online. Banyaknya jumlah produk disertai dengan banyaknya dengan toko online yang menawarkan berbagai macam produk-produk dalam jumlah besar sehingga menimbulkan kebingungan serta keraguan dalam melakukan kegiatan belanja online. Hal yang biasa dilakukan oleh user atau konsumen dalam melakukan kegiatan belanja terkadang cenderung membuka beberapa situs belanja secara bersamaan untuk mencari produk yang sama, sedangkan banyak sekali situs belanja yang tersedia diluar sana sehingga timbul keraguan dalam memilih produk sehingga waktu cenderung dihabiskan hanya untuk melakukan pencarian dan terkadang tidak menemukan produk yang sesuai. Untuk mengatasi hal tersebut maka dibangunlah Sistem Pencarian Produk dengan mengimplementasikan *Webscraping* dengan menggunakan metode XPath dalam melakukan ekstraksi data. Data dari berbagai macam situs belanja dapat kita rekonstruksi ulang kedalam sistem kita serta dari hasil pengujian bahwa dengan memanfaatkan *Webscraping* data yang diambil akan lebih cepat serta proses pencarian akan semakin efisien.

KATA PENGANTAR

Segala puji dan syukur kehadiran Allah SWT atas berkah, rahmat dan hidayah-Nya yang senantiasa dilimpahkan kepada penulis, sehingga bisa menyelesaikan skripsi dengan judul “IMPLEMENTASI *WEBSCRAPING* UNTUK PENCARIAN PRODUK DENGAN METODE *EXTENSIBLE PATH LANGUAGE* (XPATH) PADA HALAMAN WEBSITE *E-COMMERCE*” sebagai syarat untuk menyelesaikan Program Sarjana (S-1) pada Program Sarjana Sistem Informasi STMIK PPKIA Pradnya Paramita.

Dalam penyusunan skripsi ini banyak hambatan serta rintangan yang penulis hadapi namun pada akhirnya dapat melaluinya berkat adanya bimbingan dan bantuan dari berbagai pihak baik secara moral maupun spiritual. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Dosen wali dan dosen pembimbing yang telah memberikan dukungan pengarahan selama masa pengerjaan skripsi.
2. Kedua orang tua yang telah memberikan doa dan dukungan selama proses pembuatan skripsi.
3. Teman-teman yang selalu memberikan dukungan selama pengerjaan skripsi.

Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu memberikan dukungan. Penulis mohon maaf atas segala kesalahan yang pernah dilakukan. Semoga skripsi ini dapat memberikan manfaat untuk mendorong penelitian-penelitian selanjutnya.

Malang, 25 Juni 2020,

Mahasiswa,

Rifqi Rosyidi

16.51.0010

DAFTAR ISI

HALAMAN JUDUL	1
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS SKRIPSI	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	4
1.4. Batasan Masalah	5
1.5. Kontribusi Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1. Web Scraping	6
2.2. Django	6
2.3. HTML	7
2.4. XML	8
2.5. XPATH	8
2.5.1. Implementasi XPath	9
2.5.2. Sintaks dan Semantik XPath	9
2.6. Robots.txt	17
2.7. UML (Unified Model Language)	17
2.7.1. Use Case Diagram	18
2.7.2. Class Diagram	19
2.7.3. Activity Diagram	21

2.7.4. Sequence Diagram	22
BAB III METODOLOGI PENELITIAN	24
3.1. Analisis Permasalahan	24
3.2. Solusi yang Diusulkan	25
3.2.1. Menentukan Target Scraping Website <i>E-Commerce</i>	27
3.2.2. Menganalisis Kompatibilitas Website <i>E-Commerce</i>	27
3.2.3. Menganalisis Pola Url Target Website <i>E-Commerce</i>	33
3.2.4. Menganalisis Sturktur DOM Website <i>E-Commerce</i>	34
3.2.5. Menentukan Data yang Akan Dilakukan Proses Ekstraksi	39
3.2.6. Menentukan Pola dan Ekspresi XPath	47
3.2.7. Ekstraksi Data	49
3.2.8. Cleaning Data	50
3.2.9. Simpan Data	50
3.2.10. Penggabungan dan Rekonstruksi Data	51
3.2.11. Output, Display Data	51
3.2.12. Use Case Diagram	51
3.2.13. Class Diagram	53
3.2.14. Activity Diagram	53
3.2.15. Sequence Diagram	58
3.2.16. Perancangan Antarmuka Sistem (<i>Interface System</i>)	62
3.2.16.1. Rancangan Desain Halaman Home	62
3.2.16.2. Rancangan Desain Halaman Search	63
3.2.16.3. Rancangan Desain Halaman Flashsale	64
3.2.16.4. Rancangan Desain Halaman About Us	65
BAB IV PENGUJIAN DAN HASIL	66
4.1. Implementasi Antarmuka	66
4.1.1. Halaman Beranda	66
4.1.2. Halaman Search	67
4.1.3. Halaman Flashsale	69
4.1.4. Halaman About Us	70
4.2. Pengujian Sistem	71
4.2.1. Pengujian Halaman Beranda	71

4.2.2. Pengujian Halaman Search	72
4.2.3. Pengujian Halaman Flashsale	73
4.2.4. Pengujian Halaman About Us	74
4.3. Hasil Perbandingan	74
4.3.1. Perbandingan Struktur	75
4.3.2. Perbandingan Kecepatan	76
BAB V KESIMPULAN DAN SARAN	79
5.1. Kesimpulan	79
5.2. Saran	80
DAFTAR PUSTAKA	

DAFTAR GAMBAR

No	Nama Gambar	Halaman
2. 1	Basic Architecture Of Web Scraping	6
2. 2	Logo Django	6
2. 3	Contoh Struktur Sederhana HTML	7
2. 4	Contoh Strukur Sederhana XML	8
2. 5	Contoh Dokumen XML	11
2. 6	Contoh Node	11
2. 7	Contoh Atomic Value	12
2. 8	Nodes Relationship - Parent	12
2. 9	Nodes Relationship - Children	12
2. 10	Nodes Relationship - Siblings	13
2. 11	Nodes Relationship - Ancestor	13
2. 12	Nodes Relationship - Descendants	14
2. 13	Contoh Struktur XML untuk Selecting Nodes	14
2. 14	Contoh File Robots.txt	17
3. 1	Model Sistem Yang Diusulkan	26
3. 2	File Robots.txt Website Bukalapak	28
3. 3	File Robots.txt Website Tokopedia	29
3. 4	File Robots.txt Website Elevenia	30
3. 5	File Robots.txt Website Shopee	30
3. 6	File Robots.txt Website OLX	31
3. 7	File Robots.txt Website Uqishop	31
3. 8	File Robots.txt Website Zalora	32
3. 9	File Robots.txt Website Sorabel	32
Gambar 3. 10	Tampilan Antarmuka Pencarian Produk Halaman Elevenia	34
Gambar 3. 11	Potongan Struktur HTML Pencarian Produk Halaman Elevenia	35
Gambar 3. 12	Tampilan Antarmuka Pencarian Produk Halaman OLX	35
Gambar 3. 13	Potongan Struktur HTML Pencarian Produk Halaman OLX ..	36
Gambar 3. 14	Tampilan Antarmuka Pencarian Produk Halaman Uqishop	37
Gambar 3. 15	Potongan Struktur HTML Pencarian Produk Halaman Uqishop ..	37

No	Nama Gambar	Halaman
Gambar 3. 16	Tampilan Antarmuka Pencarian Produk Halaman Sorabel	38
Gambar 3. 17	Potongan Struktur HTML Pencarian Produk Halaman Sorabel	39
Gambar 3. 18	Tag dan Data Gambar Produk Halaman Elevenia	40
Gambar 3. 19	Tag dan Data Nama Produk Halaman Elevenia	40
Gambar 3. 20	Tag dan Data Harga Produk Halaman Elevenia	41
Gambar 3. 21	Tag dan Data Asal Produk Halaman Elevenia	41
Gambar 3. 22	Tag dan Data Link Produk Halaman Elevenia	41
Gambar 3. 23	Tag dan Data Gambar Produk Halaman OLX	42
Gambar 3. 24	Tag dan Data Nama Produk Halaman OLX	42
Gambar 3. 25	Tag dan Data Harga Produk Halaman OLX	42
Gambar 3. 26	Tag dan Data Asal Produk Halaman OLX	43
Gambar 3. 27	Tag dan Data Link Produk Halaman OLX	43
Gambar 3. 28	Tag dan Data Gambar Produk Halaman Uqishop	43
Gambar 3. 29	Tag dan Data Nama Produk Halaman Uqishop	44
Gambar 3. 30	Tag dan Data Harga Produk Halaman Uqishop	44
Gambar 3. 31	Tag dan Data Link Produk Halaman Uqishop	45
Gambar 3. 32	Tag dan Data Gambar Produk Halaman Sorabel	45
Gambar 3. 33	Tag dan Data Nama Produk Halaman Sorabel	46
Gambar 3. 34	Tag dan Data Harga Produk Halaman Sorabel	46
Gambar 3. 35	Tag dan Data Link Produk Halaman Sorabel	47
Gambar 3. 36	Use Case Diagram Sistem Pencarian	51
Gambar 3. 37	Class Diagram Sistem Pencarian	53
Gambar 3. 38	Sequence Diagram Proses Pencarian	58
Gambar 3. 39	Sequence Diagram Proses Sorting Produk	59
Gambar 3. 40	Sequence Diagram Proses Filtering	60
Gambar 3. 41	Sequence Diagram Melihat Data Flashsale	61
Gambar 3. 38	Rancangan Desain Halaman Beranda (<i>Home</i>)	62
Gambar 3. 39	Rancangan Desain Halaman <i>Search</i>	63
Gambar 3. 40	Rancangan Desain Halaman <i>Flashsale</i>	64
Gambar 3. 41	Rancangan Desain Halaman About Us	65
Gambar 4. 1	Antarmuka Halaman Beranda	67

No	Nama Gambar	Halaman
	Gambar 4. 2 Antarmuka Halaman <i>Search</i>	68
	Gambar 4. 3 Halaman Flashsale	69
	Gambar 4. 4 Halaman About Us	70

DAFTAR TABEL

No	Nama Tabel	Halaman
2. 1	Xpath Axis	9
2. 2	Sintaks Nodes XPath.	15
2. 3	Nol atau Lebih Predikat	15
2. 4	Memilih Beberapa Path	16
2. 5	Tabel Use Case	18
2. 6	Tabel Class Diagram.	19
2. 7	Tabel Activity Diagram.	21
2. 8	Tabel Sequence Diagram	22
3. 1	Pola URL Website E-Commerce Dengan Keyword Pencarian	33
3. 2	Pola dan Ekspresi XPath Halaman Elevenia	47
3. 3	Pola dan Ekspresi XPath Halaman OLX	48
3. 4	Pola dan Ekspresi XPath Halaman Uqishop	48
3. 5	Pola dan Ekspresi XPath Halaman Sorabel	49
3. 6	Activity Diagram Proses Pencarian.....	54
3. 7	Activity Diagram Proses Sorting Produk.....	55
3. 8	Activity Diagram Proses Filtering.....	56
4. 1	Tabel Pengujian Halaman Beranda.....	71
4. 2	Tabel Pengujian Halaman Search.....	72
4. 3	Tabel Pengujian Halaman Flashsale.....	73
4. 4	Tabel Pengujian Halaman About Us.....	74
4. 5	Perbandingan Struktur Produk Sebelum Ekstraksi.....	75
4. 6	Perbandingan Struktur Produk Setelah Ekstraksi.....	75
4. 7	Tabel Perbandingan Kecepatan.....	76

DAFTAR LAMPIRAN

No	Nama Lampiran
----	---------------

- | | |
|----|-----------------|
| 1. | Berita Acara |
| 2. | Riwayat Hidup |
| 3. | Listing Program |