

L A P O R A N
PENELITIAN



Implementasi Analisa Semantik Pada Webdriver Untuk
Otomatisasi Scraping Profil Pengguna Facebook

Disusun oleh:

Ketua Tim : Maulidiansyah, M.Kom

NIDN. 0723079101

Anggota : Zainal Arifin, M.Kom

NIDN. 0730038602

Lembaga Penerbitan, Penelitian, dan
Pengabdian Kepada Masyarakat (LP3M)
Universitas Nurul Jadid
Paiton Probolinggo
Tahun 2020



YAYASAN NURUL JADID PAITON

**LEMBAGA PENERBITAN, PENELITIAN, &
PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS NURUL JADID
PROBOLINGGO JAWA TIMUR**

PP. Nurul Jadid
Karanganyar Paiton
Probolinggo 67291
☎ 0888-3077-077
e: lp3m@unuja.ac.id
w: <https://lp3m.unuja.ac.id>

SURAT TUGAS

Nomor: NJ-T06/LP3M/0027/A.1/03.2020

Assalamualaikum Wr. Wb.

Yang bertanda tangan di bawah ini

Nama : ACHMAD FAWAID, M.A., M.A.
NIDN : 2123098702
Jabatan : Kepala LP3M
Nama PT : Universitas Nurul Jadid
Alamat PT : PO BOX 1 Karanganyar Paiton Probolinggo 67291

Menerangkan bahwa:

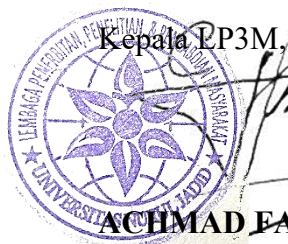
N a m a : MAULIDIANSYAH, M.KOM
NIDN : 0723079101
Jabatan : Dosen Tetap Universitas Nurul Jadid
Prodi : Teknologi Informasi
Fakultas : Teknik

Diberi tanggung jawab bersama mahasiswa sebagaimana terlampir untuk melakukan Penelitian dengan judul **“Implementasi Analisa Semantik Pada Webdriver Untuk Otomatisasi Scraping Profil Pengguna Facebook”** pada tanggal 15 Maret s.d. 30 Desember 2020

Demikian Surat Tugas ini dibuat untuk digunakan sebagaimana mestinya.

Wassalamualaikum Wr. Wb.

Paiton, 15 Maret 2020



Kepala LP3M,

ACHMAD FAWAID, M.A., M.A.

NIDN. 212309870

Lampiran Nomor: NJ-T06/LP3M/0027/A.1/03.2020

Daftar Anggota Pelaksana Penelitian
Universitas Nurul Jadid Tahun 2020

NO	NIDN/NIM	NAMA	FAKULTAS	JURUSAN
1	0723079101	Maulidiansyah, M.Kom	Teknik	Teknologi Informasi
2	0730038602	Zainal Arifin, M.Kom	Teknik	Teknologi Informasi

Paiton, 15 Maret 2020



Kepala LP3M,

ACHMAD FAWAID, M.A., M.A.
NIDN. 21230987

HALAMAN PENGESAHAN

1	Judul	:	Implementasi Analisa Semantik Pada Webdriver Untuk Otomatisasi Scraping Profil Pengguna Facebook
2	Ketua Tim	:	Maulidiansyah, M.Kom
	a. NIDN	:	0723079101
	b. Program Studi	:	Teknologi Informasi
	c. Alamat Email	:	maulid@unuja.com
3	Anggota 1	:	Zainal Arifin, M.Kom
	a. NIDN / NIM	:	0730038602
	b. Program Studi	:	Teknologi Informasi
4	Lokasi Mitra (jika ada)	:	
	a. Kabupaten	:	
	b. Provinsi	:	
5	Luaran yang Dihasilkan	:	a. Jurnal
			b.
			c.

Probolinggo, 25 Desember 2020

Mengetahui,
Kepala LP3M,

Ketua Tim,

ACHMAD FAWAID, M.A., M.A.
NIDN. 2123098702

MAULIDIANSYAH, M.KOM
NIDN. 0723079101

[Implementasi Analisa Semantik Pada Webdriver Untuk Otomatisasi Scraping Profil Pengguna Facebook]

Abstrak.

Penelitian ini membahas tentang implementasi Analisa semantik pada webdriver untuk mendapatkan data profil pengguna facebook secara otomatis. Pada era digital ini, data adalah sesuatu yang sangat berharga. Hal ini dikarenakan data dapat di olah menjadi pengetahuan baru yang bisa digunakan dalam semua lini kehidupan seperti pendidikan, ekonomi, bisnis dan lain-lain. Facebook adalah salah satu tempat penyimpanan data dalam jumlah besar. Hanya saja tidak mudah mendapatkan data dari Facebook. Dibutuhkan sebuah teknik yang bisa digunakan oleh sebuah bot untuk bisa menjelajah halaman Facebook, mengambil data profil pengguna dan mengelompokkannya berdasarkan kategori profil. Teknik tersebut adalah membuat bot dari Webdriver dengan mengimplementasikan analisa semantik.

Tahapan proses dalam penelitian ini adalah : Memahami struktur HTML pada halaman Facebook untuk mengetahui pola data profil pada Facebook. Kemudian Bot akan dikembangkan agar bisa membuka halaman profil pengguna, mengambil data profil dan kemudian menganalisanya berdasarkan pola pada struktur HTML halaman Facebook. Berdasarkan hasil uji coba pada internet dengan kecepatan 20Mbps, bot ini dapat mendapatkan 100 data profil pengguna Facebook dalam waktu \pm 20 menit.

Katakunci: Analisa Semantik, Webdriver, Facebook Bot

BAB I PENDAHULUAN

Keberadaan data menjadi sangat berharga pada era ini. Hal ini karena dengan pengolahan data dapat diperoleh pengetahuan baru yang bisa digunakan untuk banyak hal seperti penelitian, memprediksi harga saham, melihat ketertarikan pembeli, analisa sentimen dan lain-lain. Proses tersebut biasa disebut dengan data mining. Data mining adalah proses untuk mengubah sesuatu yang tidak bermanfaat menjadi sesuatu yang berharga dan dapat memberikan suatu informasi yang penting[1]. Salah satu tempat yang menyimpan data penting dalam jumlah besar adalah sosial media seperti Facebook. Tapi tidak mudah untuk mendapatkan dari situs tersebut. Karena facebook tidak mengizinkan bot penjelajah otomatis untuk mengakses halamannya. Maka dari itu butuh tool seperti webdriver untuk mendapatkan data pengguna Facebook secara otomatis dan analisa semantik untuk mengolah dan mengelompokkan data pengguna pada kategori profil seperti pendidikan, alamat, link media sosial, no telepon, whatsapps dan lain-lain. Jadi dengan teknik ini pengambilan data pengguna dari situs Facebook bisa lebih cepat dan akurat.

Semantik bertujuan untuk menerjemahkan kalimat menjadi bentuk representasi lain[2]. Sedangkan teknik yang digunakan untuk mengekstraksi data HTML pada halaman Facebook adalah Regular Expression. Regular Expression berfungsi adalah teknik yang dominan digunakan untuk mengekstrak informasi saat ini[3].

A. Kajian teori

1. Web Semantik Dalam dunia teknologi informasi penggunaan Word Wide Web (WWW) sangat banyak digunakan untuk mencari informasi berdasarkan metadata, dalam pencarian metadata tersebut lebih dikenal dengan WEB Semantik dimana metode tersebut adalah sebuah kumpulan teknologi dalam dunia komputer yang mampu membedakan sebuah informasi yang berlandaskan pada metadata. Hal ini diharapkan dapat membantu pengguna dalam melakukan pencarian informasi secara cepat dan akurat[4].
2. Arsitektur Web Semantik
 - 1) RDF (Resource Description Framework) Sumber informasi yang didapatkan dan sebuah web baik itu biodata, tempat maupun sebuah dokumen dan data lainnya sehingga penggunaan URI (Uniform Resource Identifier) sebagai alat untuk mengidentifikasi secara unik dalam membedakan sebuah resource. Agar penggunaan sumber daya sebuah web (Web Resource) mampu dideskripsikan secara jelas, maka direkomendasikanlah sebuah bahasa RDF (Resource Description Framework) oleh lembaga Word Wide Web Consortium / W3C sebagai standart dalam mendeskripsikan sebuah web resource [2].

- 2) OWL (Web Ontology Language) OWL merupakan salah satu bahasa yang telah direkomendasi oleh lembaga W3C, dimana bahasa tersebut merupakan bahasa ontologi yang paling ekspresif digunakan dalam sebuah aplikasi semantik web yang digunakan untuk merepresentasi sebuah arti dari sebuah terminologi-terminologi pada sebuah pengetahuan dimana relasi pada terminologi tersebut tidak bisa dipisah dari dokumen dan aplikasi web. Penggunaan OWL dimaksudkan untuk melakukan sebuah representasi makna dari sebuah kosakata serta sebuah relasi antara kata dengan kata yang lain sehingga dapat menghasilkan sebuah informasi yang eksplisit. Pembuatan OWL dibentuk dari penggabungan RDF dan RDFS dan ditambahkan vocabulary agar sebuah classes, properties serta individuals dapat didefinisikan. OWL memberikan pilihan sub bahasa yang dapat digunakan sesuai dengan kebutuhan, dimana pilihan tersebut adalah OWL Lite, OWL DL, serta OWL FULL[5].
- 3) SPARQL (SPARQL Protocol and RDF Query Language) Selain RDF dan OWL W3C juga membuat sebuah standart tentang penggunaan query dalam memperoleh data dari web resource, dimana standart tersebut dinamakan SPARQL (SPARQL Protocol and RDF Query Language). SPARQL Query memiliki triple pattern yaitu Subjek, predikat serta objek dimana dari masing-masing kriteria tersebut dapat dijadikan sebuah variabel dalam penggunaan SPARQL. Penggunaan Query SPARQL berdasarkan kepada proses mencocokkan sebuah pola triple pada RDF, ada beberapa syntax yang dapat digunakan pada query SPARQL diantaranya adalah:
 - a) PREFIX Statement PREFIX Adalah sebuah cara menunjukkan informasi yang ada dalam sebuah resource dimana cara ini diwakili oleh URI (Uniform Resource Identifier), pada intinya PREFIX ini digunakan untuk membuat simple sebuah sumber data.
 - b) SELECT Statement SELECT Pendefinisian sebuah variabel dimana hasilnya akan dikembalikan sebagai hasil proses query, Setiap variabel diawali dengan notasi "?".
 - c) WHERE Penggunaan WHERE digunakan untuk mendefinisikan sejumlah pattern yang wajib dimiliki dari setiap proses query. Semua pola yang mewakili sebuah kalimat RDF harus sesuai dengan RDF Triple yang terdiri dari sebuah subjek, predikat, dan objek. RDF triple tersebut dapat direpresentasikan oleh URI atau sebuah variable dan nilai literal.

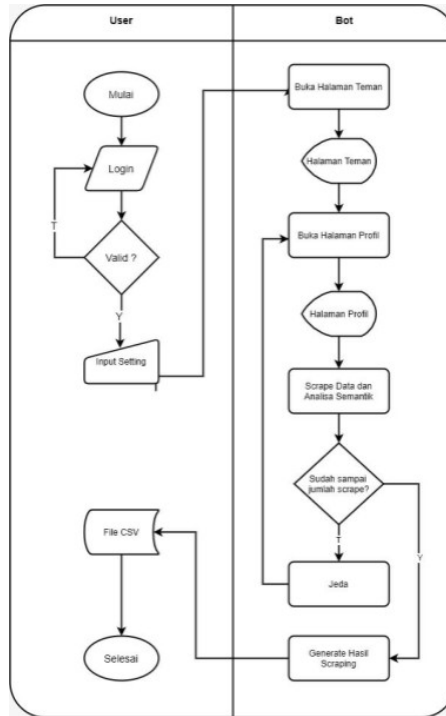
memberikan masukan sesuai dengan hasil. Langkah ini dilakukan untuk mevalidasi desain dan aplikasi agar siap diterapkan.

- e. Perbaikan Desain Proses perbaikan desain dilakukan untuk menghasilkan desain yang lebih sempurna, proses ini dilakukan mengacu kepada hasil validasi desain yang telah dilakukan.
- f. Uji Coba Produk Setelah desain dan aplikasi selesai dibuat maka perlu ada proses uji coba produk untuk mengetahui kesesuaian aplikasi dengan kebutuhan, pada proses ini ketua pengusul melakukan uji coba pencarian dengan menggunakan metode semantic, sedangkan anggota pengusul melakukan dokumentasi dari hasil percobaan.
- g. Revisi Produk Revisi produk dilakukan apabila dalam tahapan uji coba mengalami ketidaksesuaian dengan sistem, revisi ini dilakukan dengan mengacu kepada hasil dokumentasi uji coba.
- h. Uji Coba Pemakaian Proses pengujian ini dilakukan dengan sistem black box testing. Menurut Nidhra dan Dondeti black box testing juga disebut functional testing, sebuah teknik pengujian fungsional yang merancang test case berdasarkan informasi dari spesifikasi[6]. Pada proses ini ketua pengusul membuat draf kuisisioner untuk diberikan kepada masyarakat, sedangkan anggota pengusul membantu ketua dalam penyebaran kuisisioner.
- i. Revisi Produk Lanjut Pada proses ini dilakukan proses perbaikan yang mengacu pada hasil kuisisioner yang telah direkap oleh anggota pengusul, yang kemudian akan dijadikan acuan oleh ketua pengusul dalam melakukan perbaikan sistem dan aplikasi.
- j. Pembuatan Produk Masal Proses ini adalah proses akhir yang dilakukan dalam proses pengembangan aplikasi agar bisa digunakan oleh masyarakat

BAB III

HASIL DAN PEMBAHASAN

Hasil Uji Coba Dan Pembahasan



Gambar 2. Flowchart Sistem

Dalam proses ekstraksi data profil pengguna Facebook, Kami menggunakan analisa semantik untuk mengenali kategori profil pengguna. Hal ini dilakukan karena data profil yang ditampilkan di halaman facebook tidak memiliki label kategori seperti data pendidikan. Semantik dapat menyamakan kata-kata atau kalimat yang memiliki makna yang sama meskipun kata-kata atau kalimat tersebut berbeda[7].

PENDIDIKAN



MTs Wali Songo I Kedungsari Maron
Angkatan 1999 · Probolinggo



IAI Nurul Jadid
Probolinggo



MANJ
Probolinggo

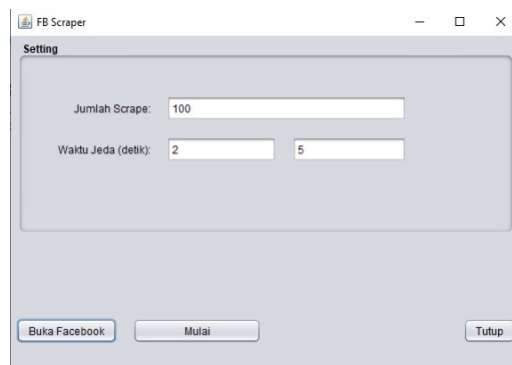


MI Tarbiyatus Syibyan Sebaung Gending Prob
Probolinggo

Gambar 3. Halaman Pendidikan Pengguna Facebook
[Sumber: facebook.com]

Pada gambar di atas facebook hanya menampilkan riwayat pendidikan pengguna, tapi tidak disebutkan jenjang pendidikan pada setiap riwayat pendidikan. Dengan semantik, kita bisa mengetahui jenjang setiap pendidikan hanya dari nama instansi pendidikan tersebut. Analisa semantik mengenali makna dari kata/kalimat berdasarkan pola. Pola dapat diketahui dengan menganalisa struktur HTML pada halaman Facebook. Contoh, jika pada jenjang pendidikan terdapat kata SMA, MA, SMK dan kata yang berkaitan dengan 3 kata tersebut seperti “Sekolah Menengah Atas”, maka Riwayat pendidikan tersebut akan diberi masukkan ke jenjang pendidikan menengah atas.

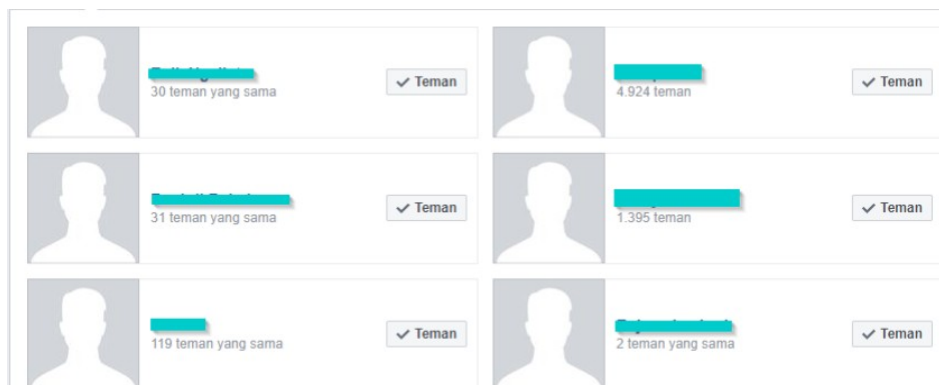
Bot dikembangkan menggunakan bahasa pemrograman Java dan plugin Selenium Webdriver. Plugin ini bisa digunakan untuk membuat otomatisasi browser yang tangguh, terukur dan dapat didistribusikan ke banyak lingkungan sistem[8]. Adapun tampilan dari Bot adalah seperti gambar 3.



Gambar 4. Tampilan Bot

Pada Bot terdapat 2 setting yang harus di isi, pertama jumlah scrape dan waktu jeda. Waktu jeda ini bersifat acak antara waktu minimum dan maksimum. Hal ini digunakan untuk mengelabui bot detector milik Facebook dengan membuat bot ini berperilaku seolah-olah manusia. Setelah itu, pengguna user bisa menekan tombol “Buka Facebook” untuk membuka halaman login Facebook dan User akan diminta login secara manual. Selanjutnya user dapat menekan tombol “Mulai” untuk memulai proses scraping. Adapun tahapan proses scraping adalah sebagai berikut:

1. Bot akan membuka halaman daftar teman



Gambar 5. Halaman Daftar Teman Facebook

2. Pada halaman daftar teman, Bot akan membuka halaman profil setiap teman facebook satu-persatu



Gambar 6. Halaman “Tentang” Facebook

3. Bot akan mengambil data profil teman facebook dan mengelompokkan data tersebut menggunakan metode Analisa Semantik (Regular Expression) Adapun hasil dari scraping bot ini adalah berupa file CSV yang berisi data profil pengguna Facebook yang sudah dikelompokkan ke dalam kolom-kolom.

Tabel 1: Data hasil scraping

Nama User	Jenis Kelamin	Riwayat Pekerjaan	Dasar	Menengah Pertama	Menengah Atas
Mahmud	Laki-laki	Guru di SMP 1 Probolinggo, Programmer di PT Jaya Makmur	-	SMP NJ	SMK NJ
Siti	Perempuan	Admin di Toko ABC, Owner TOKO XYZ	SD Negeri 1 Kraksaan	MTs NJ	MA Nurul Jadid
Rudianto	Laki-laki	Pelajar	-	MTs Miftahul Huda	SMK 2 Probolinggo
M. Fajar	Laki-laki	Setia Mencari Cinta	-	SMP NJ	SMA Nurul Jadid

Tabel 1: Data hasil scraping (Lanjutan)

Perguruan Tinggi	No Telp	Alamat	Kepercayaan	Bahasa
UNUJA	085234124865	Probolinggo	Islam	Jawa, Indonesia
UNUJA	085234125462	Surabaya	Islam	Madura, Indonesia
UNUJA	085264532569	Kraksaan	Islam moderat	Madura, Indonesia
-	082315648952	paiton	Islam	Indonesia

BAB IV PENUTUP

Berdasarkan hasil uji coba pada internet dengan kecepatan 20Mbps, Bot ini dapat mendapatkan 100 data profil pengguna Facebook dalam waktu \pm 20 menit. Bot ini juga bisa dijalankan secara paralel untuk mendapatkan hasil scraping yang lebih cepat dan singkat. Selain itu, kecepatan internet dan spesifikasi PC yang digunakan akan berpengaruh terhadap kecepatan Bot. Dari hasil uji coba tersebut dapat disimpulkan bahwa Bot ini bisa mendapatkan data profil pengguna Facebook dalam jumlah banyak dengan waktu yang relatif singkat. Selain itu, Bot ini juga bisa mengelompokkan data tersebut berdasarkan kategori profil seperti pendidikan.

DAFTAR PUSTAKA

- S. Rizal, P. Studi, T. Informatika, and U. Yudharta, "Penerapan Algoritma Naïve Bayes Untuk Prediksi Penerimaan Siswa Baru Di Smk Al- Amien Wonorejo," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 10, no. 1, pp. 14–21, 2018, doi: 10.35891/explorit.v10i1.1671.
- J. Suciadi, "Studi Analisis Metode-Metode Parsing Dan Interpretasi Semantik Pada Natural Language Processing," *J. Inform.*, vol. 2, no. 1, pp. 22–22, 2001, doi: 10.9744/informatika.2.1.pp.13-22.
- Y. Li, R. Krishnamurthy, S. Raghavan, S. Vaithyanathan, and H. V. Jagadish, "Regular expression learning for information extraction," *EMNLP 2008 - 2008 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf. A Meet. SIGDAT, a Spec. Interes. Gr. ACL*, no. October, pp. 21–30, 2008, doi: 10.3115/1613715.1613719.
- M. Unik and M. Ramli, "Penerapan Metode Semantik Search Dalam Sistem Informasi Pencarian Dokumen Kerja Praktek Dan Skripsi Berbasis Web," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 3, no. 2, pp. 83–88, 2018, doi: 10.36341/rabit.v3i2.480.
- Y. F. Badron, F. Agus, and H. R. Hatta, "STUDI TENTANG PEMODELAN ONTOLOGI WEB SEMANTIK DAN PROSPEK PENERAPAN PADA BIBLIOGRAFI ARTIKEL JURNAL ILMIAH," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 2338– 5081, 2017, [Online]. Available: <http://journal.student.uny.ac.id/ojs/ojs/index.php/pgsd/article/viewFile/135/130>.
- S. Nidhra, "Black Box and White Box Testing Techniques - A Literature Review," *Int. J. Embed. Syst. Appl.*, vol. 2, no. 2, pp. 29–50, 2012, doi: 10.5121/ijesa.2012.2204.
- N. Riza Akbar, R. Faisal, and H. Budi, "Pengembangan Data Uji Sistem Komputasi Kemiripan Teks Secara Semantik Berbahasa Indonesia," *Semin. Inform. Apl. Polinema*, 2016.
- Anonymous, "Selenium WebDriver. Internet: <https://www.selenium.dev>, 03 September 2020".