

Otomatisasi Pengujian Aplikasi Scraping Profil Pengguna Facebook Menggunakan Selenium

Muhammad Adji Assalmi^{1*}, Diki Hidayat¹, Aldiyansyah¹, Ibnu Sholih Taghalli¹, Aries Saifudin¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}adji.salmi@gmail.com, ²dicky0996@gmail.com, ³aldiyansyahwisnu@gmail.com,
⁴ibnusolihtaghalli@gmail.com, ⁵aries.saifudin@unpam.ac.id

(* : coressponding author)

Abstrak– Penelitian ini akan menunjukkan kepada Anda cara mengotomatiskan pengujian scraper Anda untuk secara otomatis menangkap informasi profil pengguna Facebook. Di era digital ini, informasi sangatlah berharga. Karena informasi dapat diolah menjadi pengetahuan baru yang dapat dibagikan di segala bidang kehidupan seperti pendidikan, bisnis dan perdagangan. Facebook adalah arsip data yang sangat besar. Mendapatkan informasi dari Facebook tidaklah mudah. Yang Anda butuhkan adalah teknik bot Anda untuk mencari halaman Facebook, mendapatkan informasi profil pengguna dan mengelompokkannya berdasarkan kategori profil. Teknik membuat bot dari Selenium Langkah-langkah proses penelitian ini antara lain: memahami struktur HTML halaman Facebook dan menemukan pola pada data profil Facebook. Bot kemudian dikembangkan yang dapat membuka halaman profil pengguna, mengambil informasi profil dan menganalisisnya berdasarkan pola struktur HTML halaman Facebook. Berdasarkan hasil pengujian internet pada kecepatan 20 Mbps, robot ini dapat mengambil 100 detail profil pengguna Facebook dalam waktu sekitar 20 menit.

Kata Kunci: Otomatisasi; Aplikasi Scraping; Facebook Bot; Selenium.

Abstract– This research describes how to automate testing scraping applications to automatically retrieve Facebook user profile data. In the digital age, data is extremely valuable. Because data can be processed into new knowledge that can be used in all areas of life such as education, business and business. Facebook is a place where a lot of data is stored. Getting data from Facebook is not easy. I need a technique that my bot can use to crawl Facebook pages, retrieve user profile data, and group it by profile category. The technique is to create a bot out of Selenium. The stages of this research process are: Understand the HTML structure of Facebook pages to find patterns in Facebook profile data. A bot is then developed that can open the user's profile page, retrieve the profile data, and parse it against the patterns in her HTML structure on the Facebook page. Experimented on the web at a speed of 20 Mbit/s, the bot was able to retrieve the details of 100 Facebook user profiles of her in 20 minutes.

Keywords: Automation; Application Scraping; Facebook Bots; Selenium.

1. PENDAHULUAN

Quality Assurance adalah serangkaian proses sistematis untuk menentukan apakah suatu produk atau jasa memenuhi syarat yang ditentukan. QA menentukan dan menetapkan persyaratan untuk membuat atau mengembangkan produk tertentu agar memiliki kualitas yang baik. Kenapa kualitas produk penting? Untuk menjaga kredibilitas perusahaan, meningkatkan kepercayaan konsumen, meningkatkan proses kerja, dan membuat perusahaan mampu bersaing dengan kompetitor. Quality Assurance menggunakan pendekatan proses untuk mencegah cacat produk. Karena itu biasanya QA juga melakukan monitoring pembuatan produk mulai dari tahap perencanaan sampai pengujian. Ini dilakukan untuk mengurangi rework (pengerjaan ulang) agar proses kerja efisien dan menghindari keluhan dari konsumen.

Pada titik ini kehadiran data menjadi sangat berharga. Pengolahan data menghasilkan pengetahuan baru yang dapat digunakan untuk penelitian, prediksi stock, identifikasi minat pembeli, analisis sentiment, dan banyak lagi. Proses ini disebut penambangan data. Data mining adalah proses mengubah sesuatu yang tidak berguna menjadi sesuatu yang berharga yang dapat memberikan informasi penting.

Salah satu tempat menyimpan banyak data penting adalah media sosial seperti Facebook. Namun mendapatkan data dari situs-situs tersebut tidaklah mudah. Ini karena Facebook tidak mengizinkan akses ke halaman dengan melihat bot secara otomatis.

Jadi Anda memerlukan aplikasi scraping yang secara otomatis menangkap data dan analitik pengguna Facebook, memproses data, dan mengklasifikasikan data pengguna ke dalam kategori profil seperti pendidikan, alamat, tautan media sosial, nomor telepon, whatsapp, dan lain sebagainya. Oleh karena itu, dengan menggunakan Teknik ini, anda dapat mengambil data pengguna dari halaman Facebook dengan lebih cepat dan akurat. Analisis ini bertujuan untuk mengubah kalimat menjadi bentuk verbal lain.

Teknik yang digunakan untuk mengekstrak data HTML di halaman Facebook adalah ekspresi reguler. Fungsi ekspresi reguler adalah teknik utama yang digunakan untuk mengekstrak informasi terkini.

2. METODOLOGI PENELITIAN

Metodologi yang digunakan dalam pengujian ini adalah pengujian dengan menggunakan selenium, yaitu:

a. Potensi dan Masalah

Untuk memahami masalah yang Anda miliki, Anda perlu melakukan beberapa penelitian pendahuluan. Hal ini dilakukan untuk memperoleh informasi dan mengumpulkan informasi, serta mencari referensi yang relevan untuk mendukung penelitian.

b. Mengumpulkan Informasi

Proses pengumpulan data dilakukan dengan melakukan eksperimen pengumpulan data tanpa menggunakan metode selenium. Tugas utama pelamar pada tahap ini adalah memperoleh informasi tentang penggunaan selenium.

c. Desain Produk

Proses ini bekerja setelah proses pendataan selesai. Hal ini dilakukan untuk mempermudah pengembangan program. Dalam proses ini, pemrakarsa utama merancang sistem dan anggota umpan balik memberikan informasi terkait hasil pencarian.

d. Validasi Desain

Untuk menentukan kecukupan desain sistem, desain harus diuji dalam aplikasi akhir. Pemrakarsa utama menyajikan hasil proyek dan rancangan aplikasi, dan anggota pemimpin memberikan umpan balik berdasarkan hasil. Langkah ini dilakukan untuk memvalidasi desain dan program sehingga siap untuk diimplementasikan.

e. Perbaiki Desain

Berdasarkan hasil pengujian desain, kami menerapkan proses penyempurnaan desain untuk membuat desain yang lebih lengkap.

f. Uji Coba Produk

Setelah desain dan aplikasi dibuat, harus mengikuti proses pengujian produk untuk menentukan apakah memenuhi persyaratan aplikasi. Dalam proses ini, pemrakarsa utama melakukan percobaan pencarian menggunakan metode Selenium, dan anggota pendukung mendokumentasikan hasil percobaan.

g. Revisi Produk

Setiap penyimpangan dari sistem selama fase pengujian akan menghasilkan tinjauan produk. Review ini didasarkan pada hasil kertas ujian.

h. Uji Coba Pemakaian

Proses pengujian ini dilakukan dengan menggunakan sistem pengujian black box. Menurut Nidhra dan Dondeti, pengujian black box, juga dikenal sebagai pengujian fungsional, adalah metode pengujian fungsional yang merancang kasus uji berdasarkan informasi spesifikasi. Dalam proses ini, pengusul utama menyiapkan kuesioner untuk dibagikan kepada masyarakat, dan anggota pengusul membantu presiden dalam menyebarkan kuesioner.

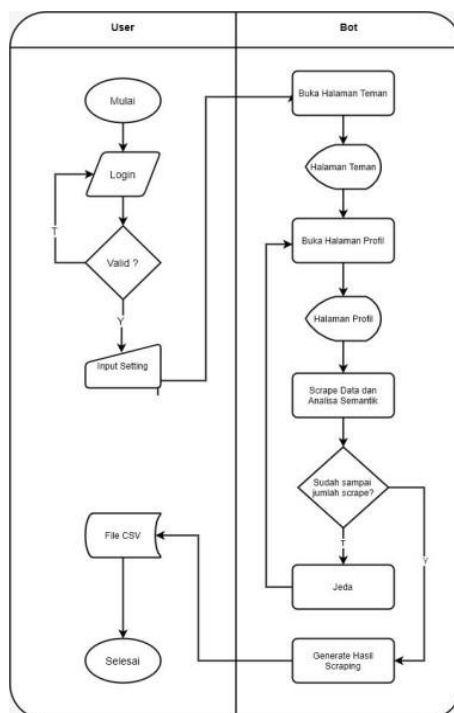
i. Revisi Produk Lanjut

Pada titik ini, hasil survei, yang dirangkum oleh anggota proposal, dibagikan dan proses perbaikan dilakukan sebagai referensi untuk perbaikan sistem dan implementasi dari pengusul utama.

j. Pembuatan Produk Masal

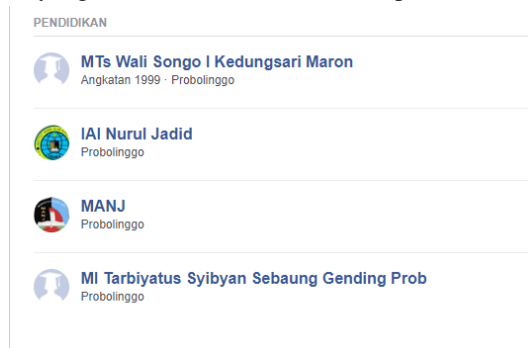
Proses ini merupakan langkah terakhir dalam proses pengembangan aplikasi agar tersedia untuk umum.

3. HASIL DAN PEMBAHASAN



Gambar 1. Flowchart Sistem

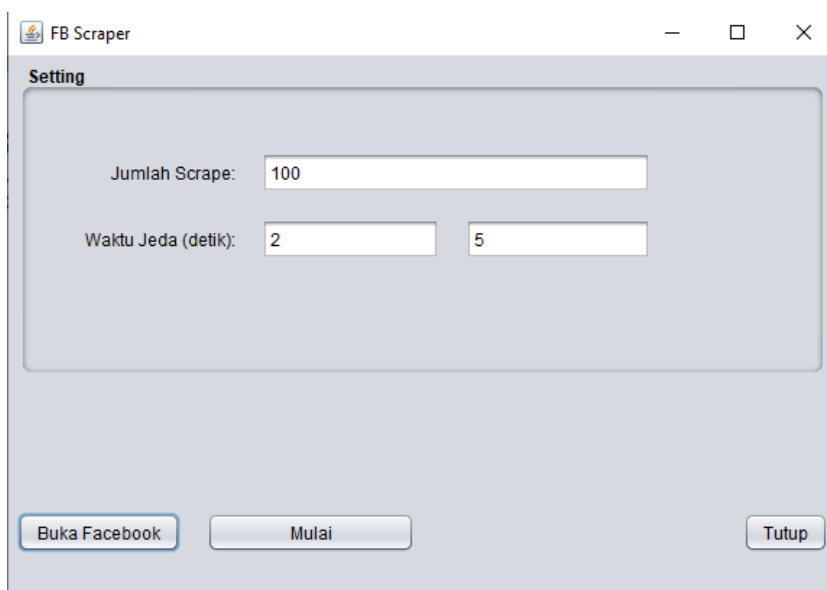
Saat mengekstrak data profil pengguna Facebook, memakai analisis semantik untuk mengidentifikasi kategori profil pengguna. Hal ini dikarenakan data profil yang ditampilkan pada page Facebook tidak mempunyai label kategori misalnya data Pendidikan. Semantik memungkinkan istilah kata dan frasa dengan makna yang sama buat disamakan meskipun istilah kata dan frasa berbeda.



Gambar 2. Halaman Pendidikan Pengguna Facebook

Pada gambar di atas, Facebook hanya menampilkan tingkat pendidikan pengguna, tetapi setiap tingkat pendidikan tidak menyebutkan pendidikan. Dengan semantik, Anda dapat mengetahui tingkat pendidikan hanya dengan nama institusi. Analisis semantik mengenali makna kata/frasa berdasarkan pola. Pola dapat diidentifikasi dengan menganalisis struktur HTML halaman Facebook. Misalnya, jika kata SMA, MA, SMK dan kata-kata yang terkait dengan ketiga kata tersebut, seperti "Gymnasium", muncul di tingkat pendidikan, maka sejarah pendidikan dimasukkan di tingkat pendidikan tinggi.

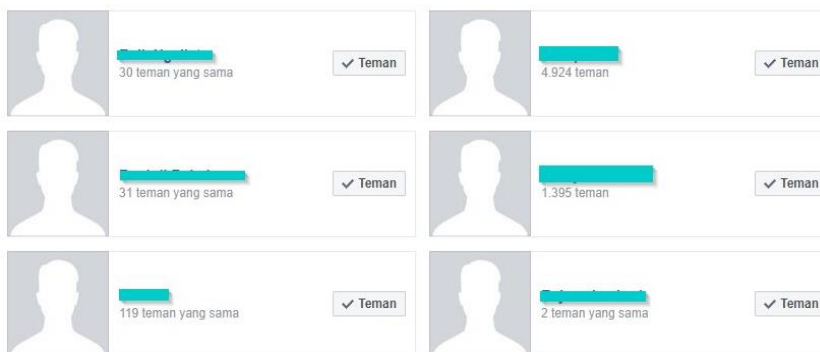
Bot ini dikembangkan memakai bahasa pemrograman Java dan plugin Selenium Web Driver. Plugin ini memungkinkan Anda menciptakan otomatisasi browser yang kuat, skalabel, dan bisa didistribusikan pada macam lingkungan sistem. Gambar 2 menunjukkan seperti apa bot itu.



Gambar 3. Tampilan Bot

Bot memiliki 2 pengaturan yang wajib Anda masukkan. Yang pertama merupakan jumlah gesekan dan waktu jeda. Waktu tunda ini rambang antara waktu saat minimum dan maksimum. Ini dipakai untuk mengelabui detektor bot Facebook supaya berperilaku misalnya manusia. Pengguna lalu bisa mengklik tombol Buka Facebook buat membuka laman login Facebook dan pengguna akan buat login secara manual. Pengguna lalu bisa menekan tombol "Start" untuk memulai proses scraping. Tahapan menggores merupakan:

1. Bot membuka halaman daftar teman



Gambar 4. Halaman Daftar Teman Facebook

- Di halaman daftar teman, bot membuka halaman profil setiap teman Facebook secara bergantian.



Gambar 5. Halaman “Tentang” Facebook

- Bot akan mengambil data profil teman facebook dan mengelompokkan data tadi menggunakan metode Analisa Semantik (Regular Expression)

Hasil dari bot scraping ini berupa file CSV dengan data profil pengguna Facebook yang dikelompokkan ke dalam 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 2. 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	paition	Islam	Indonesia

4. KESIMPULAN

4.1 Kesimpulan

Bot diuji di web dengan kecepatan 20Mbps, dan mampu mengambil detail 100 profil pengguna Facebook dalam waktu kurang lebih 20 menit. Bot ini juga dapat berjalan secara paralel untuk hasil scraping yang lebih cepat dan lebih pendek. Selain itu, kecepatan internet dan spesifikasi PC Anda juga mempengaruhi kecepatan bot Anda. Dari hasil pengujian, kami dapat menyimpulkan bahwa bot ini dapat menerima sejumlah besar data profil dari pengguna Facebook dalam waktu yang relatif singkat. Selain itu, bot juga dapat mengelompokkan data berdasarkan kategori profil seperti pendidikan.

4.2 Saran

Dengan dilakukannya pengujian ini, kita bisa mendapatkan bahan untuk evaluasi terhadap aplikasi ini untuk meminimalisir bugs atau error yang ada pada aplikasi ini, dan untuk pengembangan ke tahap yang lebih baik lagi.

REFERENCES

- A. Yunistira, D. H. (2020). Analisis Penerapan model business intelligence pada Aplikasi payment point online banking dalam Meningkatkan Strategi Pemasaran (Studi Kasus: Aplikasi ApotikKuota). *Jurnal Ilmu Komputer dan Agri-Informatika*, 7(1), 1-10. doi:10.29244/jika.7.1.1-10
- F. Firman, F. F. (2021). Aplikasi Pengingat Dan Pendataan Kenaikan Golongan Gaji Berbasis web Menggunakan Metode white box testing Dan black box testing. *Jurnal Teknologi Informasi*, 7(1), 50-57. doi:10.52643/jti.v7i1.1387
- I. R. Bakti, Y. P. (2021). Rancang bangun sistem informasi geografis (Sig) lokasi praktek kerja industri (Prakerin) smk Methodist Medan berbasis web. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 6(1), 1-6. doi:10.36341/rabit.v6i1.1505
- Jollyta, D. Z. (2020). Penerapan Metode left corner parsing Dan Analisis Kontekstual Pada natural language processing. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 7(1), 21-28. doi:10.35957/jatisi.v7i1.269
- N. Izzah, N. Y. (2022). Sistem Deteksi Kemiripan Teks Pada Berita Berbahasa Indonesia Menggunakan algoritma Ratcliff/Obershelp. *Jurnal Linguistik Komputasional (JLK)*, 5(1), 1. doi:10.26418/jlk.v5i1.65
- R. Mande, K. C. (2018). Regular expression rule-based algorithm for multiple documents key information extraction. 2018 *International Conference on Smart Systems and Inventive Technology (ICSSIT)*. doi:10.1109/icssit.2018.8748764
- Raghavendra, S. (2020). *Introduction to selenium. Python Testing with Selenium*, 1-14. doi:10.1007/978-1-4842-6249-8_1
- S. Rizal, P. S. (2018). Penerapan ALGORITMA naive Bayes untuk PREDIKSI PENERIMAAN SISWA BARU Di SMK al-amien WONOREJO. *Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, 10(1), 14-21. doi:10.35891/explorit.v10i1.1671
- W. Wahyono, A. N. (2018). Penerapan algoritma c4.5 untuk prediksi tingkat kompetensi karyawan pt multistrada arah sarana. *Joutica*, 3(1), 145. doi:10.30736/jti.v3i1.201
- Wita, E. (2021). Penerapan natural language processing Untuk Mengidentifikasi Kalimat Ambigu pada Surat Kabar Menerapkan Metode shift reduce parsing. *BEES: Bulletin of Electrical and Electronics Engineering*, 2(2), 63-66. doi:10.47065/bees.v2i2.996.