

**CHROME EXTENSION UNTUK DATA GRABBER MEDIA
SOSIAL TWITTER DENGAN METODE XPATH SELECTOR**

Skripsi



oleh
ERINDA RESHA ASTANTI
71160073

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2020

CHROME EXTENSION UNTUK DATA GRABBER MEDIA SOSIAL TWITTER DENGAN METODE XPATH SELECTOR

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

ERINDA RESHA ASTANTI
71160073

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2020

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
SKRIPSI/TESIS/DISERTASI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Erinda Resha Astanti
NIM : 71160073
Program studi : Informatika
Fakultas : Teknologi Informasi
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**“CHROME EXTENSION UNTUK DATA GRABBER MEDIA SOSIAL
TWITTER DENGAN METODE XPATH SELECTOR”**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 7 Agustus 2020

Yang menyatakan



(Erinda Resha Astanti)

NIM.71160073

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

CHROME EXTENSION UNTUK DATA GRABBER MEDIA SOSIAL TWITTER DENGAN METODE XPATH SELECTOR

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 1 Juli 2020



ERINDA RESHA ASTANTI

71160073

HALAMAN PERSETUJUAN

Judul Skripsi : CHROME EXTENSION UNTUK DATA GRABBER
MEDIA SOSIAL TWITTER DENGAN METODE
XPath Selector

Nama Mahasiswa : ERINDA RESHA ASTANTI

NIM : 71160073

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2019/2020


Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 1 Juli 2020

Dosen Pembimbing I



Antonius Rachmat C., S.Kom., M.Cs.

Dosen Pembimbing II



Yuan Lukito, S.Kom., M.Cs.

HALAMAN PENGESAHAN

CHROME EXTENSION UNTUK DATA GRABBER MEDIA SOSIAL TWITTER DENGAN METODE XPATH SELECTOR

Oleh: ERINDA RESHA ASTANTI / 71160073

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 22 Juli 2020

Yogyakarta, 3 Agustus 2020
Mengesahkan,

Dewan Penguji:

1. Antomius Rachmat C., S.Kom., M.Cs.
2. Yuan Lukito, S.Kom., M.Cs.
3. Aditya Wikan Mahastama, S.Kom., M.Cs.
4. Maria Nila Anggia Rini, S.T, M.T.I



Dekan

(Restyandito, S.Kom., MSIS, Ph.D.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah melimpahkan berkat dan anugerahnya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Chrome Extension untuk Data Grabber Media Sosial Twitter dengan Metode XPath Selector”. Penelitian ini dilakukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer di Program Studi Informatika, Universitas Kristen Duta Wacana.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kesalahan dan kekurangan sehingga dapat dikatakan jauh dari sempurna, mengingat kurangnya wawasan, pengetahuan dan pengalaman penulis. Oleh karena itu penulis berharap adanya kritik dan saran yang membangun dari semua pihak demi menyempurnakan skripsi ini.

Pada kesempatan ini, penulis ingin menyampaikan terima kasih atas bantuan, dukungan, dan bimbingan yang telah diberikan selama ini, kepada :

1. Bapak, ibu, kakak dan adik keluarga tercinta yang senantiasa memberikan doa, semangat, dukungan penuh, motivasi, dan kasih sayang sehingga penyusunan skripsi dapat selesai.
2. Bapak Restyandito, S. Kom, MSIS, Ph.D selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
3. Ibu Gloria Virgina, S.Kom., MAI., Ph.D selaku Kepala Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
4. Bapak Antonius Rachmat, S.Kom., M.Cs selaku Dosen Pembimbing I dan Bapak Yuan Lukito, S.Kom., M.Cs selaku Dosen Pembimbing II yang bersedia memberikan waktu, tenaga dan pikiran dalam memberikan bimbingan, arahan, saran, serta dukungan selama penyusunan skripsi ini
5. Teman - teman dekat penulis, Charles Eka Swandi, Yohanes Chris Kurniawan, Cindy Clara Alvionita, dan Setiawan yang senantiasa memberikan dorongan, motivasi, semangat dan dukungan kepada penulis dalam proses penyusunan skripsi.

6. Semua pihak yang tidak dapat disebutkan satu persatu yang telah ikut serta dalam memberikan dukungan baik secara langsung maupun tidak langsung dalam proses penyusunan skripsi.

Akhir kata semoga ketulusan, pengorbanan, dan bantuan yang telah diberikan dari semua pihak tersebut kiranya mendapatkan berkah dan anugerah dari Tuhan yang Maha Kuasa.

©UKDW

INTISARI

CHROME EXTENSION UNTUK DATA GRABBER MEDIA SOSIAL TWITTER DENGAN METODE XPATH SELECTOR

Kebutuhan dataset sebagai sumber penelitian untuk analisis oleh data scientist terus meningkat seiring dengan semakin meningkatnya rasa keingintahuan manusia. Misalnya seperti suatu perusahaan ingin mengetahui tanggapan konsumennya tentang produk yang baru saja diluncurkan, sehingga mereka dapat menyusun langkah selanjutnya yang akan diambil oleh perusahaan tersebut. Tentu saja untuk mengetahui hal tersebut, perlu dilakukan analisis terhadap tanggapan konsumen terlebih dahulu. Analisis dilakukan menggunakan sumber data yang dapat ditemukan salah satunya pada sosial media seperti Twitter.

Penelitian ini dimaksudkan untuk menyediakan aplikasi yang dapat digunakan menjadi data *grabber* untuk mengumpulkan data dari media sosial Twitter. Aplikasi data *grabber* dirancang sebagai Chrome *Extension* dengan implementasi metode *XPath Selector*. Dengan implementasi metode *XPath Selector* aplikasi diharapkan dapat mengumpulkan data tanpa terbatas oleh batasan akses API Twitter. Sebagai gantinya dengan menerapkan metode ini, aplikasi perlu mengakses halaman web Twitter. Pemasangan aplikasi sebagai Chrome Extension memudahkan untuk menjalankan aplikasi sembari mengakses halaman web Twitter pada browser.

Aplikasi TweetGrabberERA berhasil dibangun pada penelitian ini sesuai kebutuhan yang diharapkan. Pengujian yang dirancang telah menunjukkan bahwa aplikasi data *grabber* dengan mengimplementasikan metode *XPath Selector* berjalan dengan baik dan membuktikan bahwa aplikasi dapat menghasilkan dataset yang untuk keperluan analisis oleh data *scientist*.

Kata kunci: XPath Selector, Chrome Extension, data grabber, Twitter.

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
1.6. Metode Penelitian.....	5
1.7. Sistem Penulisan.....	6
BAB II LANDASAN TEORI.....	7
2.1 Tinjauan Pustaka dan Landasan Teori.....	7
2.1.1. Tinjauan Pustaka.....	7
2.1.2. Landasan Teori.....	9
BAB III PERANCANGAN SISTEM.....	18
3.1. Spesifikasi Sistem.....	18
3.1.1. Kebutuhan Perangkat Keras.....	18
3.1.2. Kebutuhan Perangkat Lunak.....	18
3.1.3. Kebutuhan Teknologi.....	18
3.2. Perancangan Sistem.....	19
3.2.1. Arsitektur Sistem.....	19
3.2.2. Blok Diagram Sistem.....	20
3.2.3. Halaman Antarmuka.....	26

3.2.4.	Basis Data	28
3.3.	Perancangan Pengujian Sistem.....	29
3.3.1.	Pengujian Aplikasi	29
3.3.2.	Pengujian Validasi <i>Dataset</i>	35
BAB IV IMPLEMENTASI DAN ANALISIS SISTEM		38
4.1.	Implementasi Sistem	38
4.1.1.	Fitur <i>Grabbing Data</i>	39
4.1.2.	Fitur Penjadwalan.....	42
4.1.3.	Fitur Pengelolaan XPath	45
4.1.4.	Fitur Pengelolaan <i>Database</i>	48
4.2.	Analisis Sistem.....	50
4.2.1.	Hasil Pengujian Aplikasi.....	50
4.2.2.	Hasil Pengujian Validasi <i>Dataset</i>	52
BAB V KESIMPULAN DAN SARAN.....		56
5.1.	Kesimpulan.....	56
5.2.	Saran.....	57
DAFTAR PUSTAKA		58
LAMPIRAN.....		61
	Lampiran Hasil Pengujian.....	61
	Lampiran Source Code	118
	Lampiran Kartu Konsultasi	119

DAFTAR GAMBAR

Gambar 2.1 Versi dasar manifest.json file	10
Gambar 2.2 Alur kerja <i>background script</i>	11
Gambar 2.3 <i>Pop up</i> AdBlocker Ultimate <i>extension</i>	11
Gambar 2.4 Alur kerja <i>content script</i>	12
Gambar 2.5 <i>Options page</i> dari AdBlocker Ultimate <i>extension</i>	12
Gambar 2.6 Elemen yang ditunjuk oleh operator logika AND.....	14
Gambar 2.7 Elemen yang ditunjuk oleh operator logika NOT	14
Gambar 2.8 <i>Confusion matrix</i> untuk <i>multi-class</i> (Sumber: Deng dkk., 2016)	16
Gambar 2.9 Tampilan Twitter pada desktop.....	17
Gambar 3.1 Blok diagram sistem aplikasi	19
Gambar 3.2 <i>Flowchart</i> Chrome <i>Extension</i>	20
Gambar 3.3 <i>Use case</i> diagram sistem	21
Gambar 3.4 Halaman <i>pop up</i> menu utama	26
Gambar 3.5 Halaman <i>pop up</i> menu <i>setting</i>	27
Gambar 3.6 Halaman <i>pop up</i> ketika aplikasi berjalan	27
Gambar 3.7 Halaman tambahan berisi informasi aplikasi	28
Gambar 3.8 <i>CSV Schema</i>	36
Gambar 3.9 Rancangan alur kerja pada <i>tool</i> Orange sesuai <i>widget</i>	36
Gambar 4.1 Susunan <i>file</i> dalam folder.....	38
Gambar 4.2 Kolom <i>input keyword</i> untuk pengguna pada halaman <i>pop up</i> menu utama aplikasi.....	39
Gambar 4.3 Penerapan <i>method document.evaluate()</i>	41
Gambar 4.4 Penjadwalan aktif saat posisi <i>switch on</i>	42
Gambar 4.5 Penjadwalan tidak aktif saat posisi <i>switch off</i>	43
Gambar 4.6 Tombol STOP pada halaman <i>pop up</i> aplikasi yang sedang berjalan	43
Gambar 4.7 Halaman <i>pop up</i> aplikasi yang sedang berjalan tanpa menjalankan fitur penjadwalan.....	44
Gambar 4.8 Halaman <i>pop up</i> aplikasi yang sedang berjalan dengan menjalankan fitur penjadwalan.....	44

Gambar 4.9 Tombol <i>update</i> tidak dapat diklik pada halaman <i>setting pop up</i> aplikasi	46
Gambar 4.10 Tombol <i>update</i> dapat diklik pada halaman <i>setting pop up</i> aplikasi	46
Gambar 4.11 Tombol <i>restore</i> pada halaman <i>setting pop up</i> aplikasi.....	48
Gambar 4.12 <i>File</i> CSV hasil dari ekspor data	49
Gambar 4.13 Sebelum <i>database</i> TweetGrabberERA dihapus.....	49
Gambar 4.14 Sesudah <i>database</i> TweetGrabberERA dihapus.....	50
Gambar 4.15 Hasil pengujian dari CSV Validator	52
Gambar 4.16 Alur kerja widget pada <i>tool</i> Orange	53
Gambar 4.17 Hasil analisis sentimen oleh <i>tool</i> Orange yang telah diberi label ...	54
Gambar 4.18 Hasil analisis sentimen oleh para pakar	54

©UKDW

DAFTAR TABEL

Tabel 3.1 <i>Use case</i> pengambilan data oleh pengguna.....	22
Tabel 3.2 <i>Use case</i> pembaruan XPath oleh pengguna	23
Tabel 3.3 <i>Use case</i> pengembalian ke <i>default</i> XPath oleh pengguna.....	24
Tabel 3.4 <i>Use case</i> ekspor data oleh pengguna.....	24
Tabel 3.5 <i>Use case</i> hapus data oleh pengguna	25
Tabel 3.6 Struktur tabel <i>database</i> TweetGrabberERA	28
Tabel 3.7 Skenario pengujian fitur <i>grabbing</i> data	29
Tabel 3.8 Skenario pengujian fitur penjadwalan	32
Tabel 3.9 Skenario pengujian fitur pengelolaan XPath	34
Tabel 3.10 Skenario pengujian fitur pengelolaan <i>database</i>	35
Tabel 4.1 Daftar XPath yang tersimpan pada <i>local storage</i>	40
Tabel 4.2 Daftar variabel yang digunakan untuk menampung XPath	41
Tabel 4.3 Daftar XPath Absolute untuk URL https://twitter.com/search?q=Tweet%20Grabber%20(from%3Aerindaresha)%20until%3A2020-01-01	47
Tabel 4.4 Daftar XPath Absolute untuk URL https://twitter.com/search?&q=Tweet%20Grabber%20(from%3Aerindaresha).. ..	47
Tabel 4.5 Hasil perbandingan TweetGrabberERA dengan <i>Automatic Crawling Data in Twitter</i>	51
Tabel 4.6 <i>Confusion matrix</i>	55

BAB I

PENDAHULUAN

1.1. Latar Belakang

Media sosial digunakan sebagai media komunikasi dan berbagi informasi secara daring. Media sosial telah menjadi salah satu kebutuhan hidup bagi kaum muda maupun tua sehingga kepopuleran media sosial terus meningkat dari tahun ke tahun. Fitur-fitur yang ditawarkan media sosial semakin beraneka ragam dan terus berkembang agar dapat menjadi pilihan bagi penggunanya. Salah satu media sosial yang masih menjadi tren yaitu Twitter. Twitter merupakan media sosial daring yang cukup populer, hingga saat ini tercatat 330 juta pengguna aktif tiap bulannya (Kellogg, 2020). Twitter menampilkan *tweet* yaitu kiriman yang memiliki batas tidak lebih dari 280 karakter (Twitter, tanpa tahun), membuatnya terlihat sederhana sehingga mudah untuk dibaca. Kiriman pada Twitter dapat bermanfaat dan menjadi informasi yang berguna, jika dikumpulkan dan diolah terlebih dahulu. Informasi-informasi yang banyak ditemukan biasanya berupa berita, ilmu pengetahuan, opini, dan masalah personal. Agar dapat memanfaatkan informasi yang ada, perlunya sebuah aplikasi yang dapat membantu mengumpulkan informasi-informasi tersebut dalam suatu bentuk *dataset* sehingga pengolahan informasi akan menjadi lebih mudah.

Mengumpulkan informasi dari Twitter yang bisa saja tidak sedikit jumlahnya itu tentu memerlukan waktu. Data *grabber* merupakan salah satu solusi, dimana aplikasi tersebut dapat bekerja untuk mengambil data secara otomatis. Diperlukan metode yang tepat agar aplikasi dapat berjalan dengan baik dan informasi yang diambil lengkap dan sesuai. Metode yang mudah dijumpai untuk pengambilan data pada Twitter adalah melalui API Twitter seperti yang telah diterapkan pada Tweepy. Tweepy sendiri merupakan sebuah *library* di Python yang digunakan untuk mengakses API Twitter. Sayangnya untuk mengakses API Twitter, perlu melakukan pengesahan kredensial dimana pengguna akan diminta untuk memasukkan *consumer key*, *consumer secret key*, *access token key*, *access token secret key* yang hanya didapat melalui izin dari Twitter. Hal tersebut sungguh membatasi pengguna untuk dapat mengumpulkan *tweet*. Adapun menurut Rizaldi

dan Arief (2017) menyebutkan penggunaan metode *XPath Selector* untuk melakukan web *scraping* dapat menghasilkan artikel yang lebih lengkap. Metode tersebut bisa membantu mengumpulkan informasi termasuk dari Twitter dengan mengakses halaman web Twitter. Mengakses halaman web Twitter mudah dilakukan, pengguna hanya memerlukan browser untuk melakukannya.

Untuk memudahkan penggunaan aplikasi, browser Chrome saat ini telah dilengkapi dengan *add-on* yang disebut *Chrome Extension*. Dengan adanya *Chrome Extension* pengguna dapat melakukan kustomisasi terhadap browsernya sehingga akan mendapatkan pengalaman baru yang berbeda (Google, tanpa tahun). Cukup dengan memasang *Chrome Extension* pada browser Chrome, penggunaan aplikasi sambil mengakses situs web pada browser menjadi lebih sederhana. Hal ini berguna bagi aplikasi yang perlu berjalan bersamaan dengan pengaksesan suatu halaman web contohnya seperti aplikasi pengumpulan informasi pada Twitter.

Adapun masalah yang akan dihadapi ketika menggunakan metode *XPath Selector*. Struktur web seperti media sosial Twitter ini akan selalu diperbarui dalam jangka waktu tertentu. Hal tersebut jelas akan menimbulkan kendala ketika menggunakan *XPath Selector* yang bekerja dengan cara melakukan navigasi pada halaman HTML (MDN Contributors, 2019). Perlunya penyesuaian kembali agar aplikasi tetap berjalan sesuai harapan.

Berdasarkan uraian permasalahan yang ada, penulis memutuskan untuk melakukan penelitian tentang implementasi *XPath Selector* untuk pembuatan *Chrome Extension* sebagai data *grabber* media sosial Twitter. Aplikasi akan dikembangkan menggunakan teknologi web seperti HTML, JavaScript, dan CSS. Selanjutnya akan dilakukan skenario analisis pengujian terhadap *dataset* yang dihasilkan untuk melihat apakah *dataset* sudah memenuhi standar untuk dapat digunakan dan diolah. Ada pula pengujian yang akan dilakukan terhadap aplikasi untuk menunjukkan bahwa aplikasi juga dapat bekerja pada halaman web yang dinamis.

1.2. Rumusan Masalah

Berdasarkan uraian permasalahan yang sudah disampaikan pada latar belakang, penulis memperoleh rumusan masalah sebagai berikut:

1. Bagaimana implementasi XPath *Selector* pada Chrome *Extension* untuk data *grabber* media sosial Twitter?
2. Bagaimana solusi untuk mengatasi perubahan struktur HTML pada setiap versi yang akan datang agar Chrome *Extension* untuk data *grabber* media sosial Twitter dengan metode XPath *Selector* dapat tetap bekerja?
3. Apakah data *grabber* dapat menghasilkan *dataset* yang valid untuk digunakan sebagai keperluan analisis oleh data *scientist*?

1.3. Batasan Masalah

Pada penulisan tugas akhir ini, penulis menentukan beberapa batasan masalah, antara lain:

1. Aplikasi data *grabber* berupa *add-on* yang dirancang untuk browser Google Chrome.
2. Sumber data untuk aplikasi data *grabber* ini adalah media sosial Twitter.
3. Halaman web Twitter yang diakses harus menggunakan pengaturan bahasa dalam bahasa Inggris.
4. Aplikasi data *grabber* tidak melakukan pengambilan *file* gambar dan video.
5. Aplikasi mengambil *tweet* dalam bahasa Inggris yang diatur melalui fitur pencarian pada Twitter.
6. Aplikasi mengambil data yang terdapat sesuai *keyword* yang diinputkan oleh pengguna.
7. Data yang telah dikumpulkan akan dapat diekspor dalam bentuk *file* dengan format CSV sesuai struktur kolom pada *database* yang terdiri dari kolom *tweet_id* bertipe data String, *created_at* bertipe data Datetime, *name* bertipe data String, *username* bertipe data String, *text_content* bertipe data String, dan *keyword* bertipe data String.
8. Aplikasi tidak dapat melakukan penyesuaian terhadap struktur URL Twitter yang mengalami perubahan. Penyesuaian yang dilakukan aplikasi hanya untuk struktur elemen halaman web Twitter. URL yang

digunakan untuk melakukan grabbing data adalah [https://twitter.com/search?q=\[inputKeyword\]%20lang%3Aen](https://twitter.com/search?q=[inputKeyword]%20lang%3Aen) untuk mendapatkan filter pencarian berdasarkan sort by Top dan [https://twitter.com/search?q=\[inputKeyword\]%20lang%3Aen&f=live](https://twitter.com/search?q=[inputKeyword]%20lang%3Aen&f=live) untuk mendapatkan filter pencarian berdasarkan sort by Latest, dengan keterangan bahwa [inputKeyword] merupakan keyword yang diinputkan oleh pengguna.

1.4. Tujuan Penelitian

Tujuan penelitian dari penulisan tugas akhir ini adalah sebagai berikut:

1. Membuat Chrome *Extension* untuk data *grabber* media sosial Twitter dengan metode *XPath Selector*.
2. Mengetahui solusi untuk mengatasi perubahan struktur HTML pada setiap versi yang akan datang agar Chrome *Extension* untuk data *grabber* media sosial Twitter dengan metode *XPath Selector* tetap bekerja.
3. Membuktikan bahwa Chrome *Extension* untuk data *grabber* media sosial Twitter dengan metode *XPath Selector* dapat menghasilkan *dataset* yang valid untuk digunakan sebagai sumber data oleh data *scientist*.

1.5. Manfaat Penelitian

Diharapkan penelitian tugas akhir ini dapat menciptakan Chrome *Extension* yang bekerja sebagai data *grabber* untuk mengambil *tweet* pada media sosial Twitter secara otomatis dengan metode *XPath Selector* sesuai dengan *keyword* yang dimasukkan oleh pengguna dan dapat beradaptasi terhadap perubahan struktur HTML yang akan terjadi jika ada perubahan versi dari media sosial Twitter.

1.6. Metode Penelitian

Metode penelitian ini secara garis besar dibagi menjadi beberapa tahap yaitu:

1. Studi Pustaka

Studi pustaka dilakukan dengan mencari referensi tentang data *grabber*, data *scraper*, XPath, Chrome *Extension* dan penelitian serupa lainnya dari artikel, buku, jurnal, dan literatur yang mendukung.

2. Perancangan

Penulis mempersiapkan rancangan dan desain aplikasi baik dari segi *back-end* maupun *front-end*, serta mempersiapkan skenario pengujian data penelitian yang akan diterapkan pada aplikasi. Skenario tersebut juga akan disimulasikan terlebih dahulu untuk mendapatkan data awal sebagai sumber data yang akan diteliti.

3. Membangun Aplikasi

Membangun aplikasi berupa Chrome *Extension* dengan mengimplementasikan XPath *Selector* yang dapat menghasilkan *dataset* secara otomatis. Aplikasi dirancang untuk dapat melakukan pengambilan data dari halaman media sosial Twitter yang sedang dimuat dengan cara menyeleksi elemen-elemen yang ada sesuai kata kunci yang dimasukkan.

4. Pengujian

Pada tahap ini dilakukan dua pengujian yaitu terhadap aplikasi dan terhadap *dataset* yang didapat dari aplikasi yang sudah dibangun sesuai skenario. Pengujian terhadap aplikasi dilakukan untuk menunjukkan apakah aplikasi dapat bekerja dengan baik pada halaman web yang dinamis. Kemudian juga dilakukan pengujian *dataset* untuk memastikan bahwa *dataset* yang dihasilkan dapat dinyatakan valid sebagai sumber data oleh data scientist.

5. Menyusun Kesimpulan dan Evaluasi

Menyusun kesimpulan dan evaluasi penelitian berdasarkan pengujian yang telah dilakukan.

1.7. Sistem Penulisan

Pada penulisan penelitian ini, akan dibagi dalam lima bab sesuai rincian yang tertera sebagai berikut:

BAB I PENDAHULUAN merupakan gambaran umum penelitian yang berisi latar belakang, rumusan masalah, batasan penelitian, tujuan penelitian, metode penelitian, dan struktur laporan.

BAB II LANDASAN TEORI menguraikan tentang teori-teori dasar dan pendukung yang digunakan sebagai dasar dalam pembuatan penelitian. Teori - teori tersebut berupa definisi, konsep dasar, uraian kualitatif, permodelan, persamaan atau pendapat para ahli yang penulis dapat dari studi pustaka.

BAB III PERANCANGAN SISTEM membahas tentang langkah-langkah yang dilakukan dalam perancangan aplikasi yang meliputi pilihan sistem operasi, bahasa pemrograman yang digunakan, perancangan tampilan, serta perancangan cara kerja aplikasi yang dipakai di dalam penelitian, penjelasan secara rinci algoritma, *flowchart* dan arsitektur sistem, serta penguraian secara jelas cara perancangan simulasi atau perencanaan yang akan dilakukan. Pada bab ini juga menjelaskan analisis teori-teori yang digunakan.

BAB IV IMPLEMENTASI DAN ANALISIS SISTEM berisi tentang tahap-tahap implementasi aplikasi, pembahasan hasil analisis serta evaluasi kendala-kendala yang dihadapi dalam pembuatan aplikasi.

BAB V KESIMPULAN DAN SARAN berisi kesimpulan yang berupa pernyataan singkat dan tepat yang dijabarkan dari hasil analisis kegiatan implementasi dan saran-saran untuk dapat membantu dan mengembangkan kegiatan penelitian mendatang.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Aplikasi TweetGrabberERA yang dibangun sebagai Chrome *Extension* untuk data *grabber* media sosial Twitter telah berhasil mengimplementasikan XPath *Selector*. Metode XPath *Selector* bekerja untuk mendapatkan elemen HTML berisi data yang akan diambil dari halaman web Twitter. Dengan menerapkan *method* dari DOM (*Document Object Model*) yaitu *document.evaluate()* dan menggunakan properti *singleNodeValue*, XPath dapat diakses dan menerima hasil pengembalian dengan tipe elemen *node*. Dari elemen *node* yang diterima tersebut, aplikasi masih perlu melakukan penyaringan lagi agar mendapatkan data yang dibutuhkan. Penyaringan yang dimaksudkan adalah memilah data yang ingin diambil pada *node* tersebut, contohnya seperti teks, *value* dari suatu *attribute*, atau *tag* elemen.

Walaupun menggunakan metode XPath *Selector*, aplikasi ini mampu beradaptasi pada struktur HTML halaman web Twitter yang berubah-ubah dengan adanya fitur pengelolaan XPath. Fitur pengelolaan XPath memiliki kegunaan salah satunya untuk melakukan *update* XPath jika diketahui ada perubahan struktur HTML halaman web Twitter. Cara kerja *update* XPath ini dengan memanfaatkan XPath *Absolute* pada halaman yang telah disiapkan untuk mendapatkan XPath *Relative* yang baru. XPath *Relative* inilah yang digunakan oleh aplikasi untuk mendapatkan elemen HTML dari data yang akan diambil.

Dataset dalam bentuk *file* CSV yang dihasilkan oleh aplikasi tersebut memiliki 6 kolom terdiri dari *tweet_id*, *created_at*, *name*, *username*, *text_content*, dan *keyword*. Pengujian validasi *file* format CSV yang melibatkan pengecekan skema dataset yang berisi 15.000 rekam data berhasil dilakukan dan dinyatakan lolos. Dilakukan juga analisis sentimen melalui pemberian label sesuai skor yang dihasilkan oleh *tool* Orange dapat menunjukkan adanya sentimen positif, negatif, dan netral pada setiap rekam *sample* data yang diambil. Analisis sentimen tersebut dibuktikan validitasnya melalui perbandingan dengan analisis sentimen oleh pakar.

Hasil perbandingan tersebut ditemukan bahwa persentase data yang sama antara analisis sentimen oleh *tool* Orange dengan pakar mencapai 61,33%.

5.2. Saran

Pada tahap ini penulis ingin menyampaikan beberapa saran untuk dapat menunjang penelitian lebih lanjut:

1. Penelitian yang telah penulis lakukan belum membahas mengenai performa aplikasi, sehingga melakukan perbandingan aplikasi berdasarkan performa tidak dapat dilakukan. Dengan melakukan penelitian mengenai performa, diharapkan aplikasi dapat diukur tingkat efisien dan efektivitasnya.
2. Menambahkan fitur *grabbing* data berdasarkan *username* akun Twitter untuk pengembangan aplikasi selanjutnya.

©UKD W

DAFTAR PUSTAKA

- Albert, Gunadi, K., & Setyati, E. (2014). Deteksi Helm pada Pengguna Sepeda Motor dengan Metode Convolutional Neural Network.
- Cimpanu, C. (2019, August 3). *Half of all Google Chrome extensions have fewer than 16 installs*. Diambil kembali dari ZDNet: <https://www.zdnet.com/article/half-of-all-google-chrome-extensions-have-fewer-than-16-installs/>
- Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. <http://dx.doi.org/10.1016/j.ins.2016.01.033>.
- Fishman, E. (tanpa tahun). *How to create and use hashtags*. Diambil kembali dari Twitter for Business: <https://business.twitter.com/en/blog/how-to-create-and-use-hashtags.html>
- Francisco, O. E. (2019). *Teknik Web Scraping Untuk Pencarian Produk Promo Toko Online*. Yogyakarta: STMIK AKAKOM.
- Google. (tanpa tahun). *Manifest File Format*. Diambil kembali dari Chrome: <https://developer.chrome.com/extensions/manifest>
- Google. (tanpa tahun). *Overview*. Diambil kembali dari Chrome: <https://developer.chrome.com/extensions/overview>
- Google. (tanpa tahun). *What are extensions?* Diambil kembali dari Chrome: <https://developers.chrome.com/extensions>
- Gunawan, R., Rahmatulloh, A., Darmawan, I., & Firdaus, F. (2018). Comparison of Web Scraping Techniques: Regular Expression, HTML DOM and Xpath. *International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018)*.

- Hidayat, A. S. (2018, January). *Tentang Aku dan kamu (XPath)*. Diambil kembali dari MEDIUM: <https://medium.com/@achmadsyah/tentang-aku-kamu-xpath-fl5f67b98733>
- Jundt, O., & Keulen, M. v. (2013). Sample-based XPath Ranking for Web Information Extraction. *8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2013)*.
- Kellogg, K. (2020, February 3). *The 7 Biggest Social Media Sites in 2020*. Diambil kembali dari Search Engine Journal: <https://www.searchenginejournal.com/social-media/biggest-social-media-sites/>
- Kusuma, P. (2016, Desember 26). *GrabLapak versi 1.0*. Diambil kembali dari AL-MANSHURIN IT Jasa pembuatan aplikasi hybrid: <https://almanshurin.com/aplikasi-grabber-toko-online/>
- Lou, Y., Zhang, Y., & Yuan, Z. (2013). Website Information Extraction Based On DOM-Model. *Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation (ISCCCA-13)*.
- MDN Contributors. (2019, Nov 24). *XPath*. Diambil kembali dari MDN Web Docs: <https://developer.mozilla.org/en-US/docs/Web/XPath>
- Mehta, P. (2016). Introduction to Google Chrome Extensions. Dalam *Creating Google Chrome Extensions* (hal. 1-33). New York City: Apress.
- Orange, University of Ljubljana. (tanpa tahun). *Sentiment Analysis*. Diambil kembali dari Orange: <https://orange.biolab.si/widget-catalog/text-mining/sentimentanalysis/>
- Pandey, P. (2018, September 23). *Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)*. Diambil kembali dari MEDIUM: <https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f>

- Raissi, H. (2019, February 15). *Using XPath to select elements*. Diambil kembali dari Parsehub: <https://help.parsehub.com/hc/en-us/articles/220618167-Using-XPath-to-select-elements>
- Rizaldi, T., & Arief, H. (2017). Perbandingan Metode Web Scraping Menggunakan CSS Selector dan Xpath Selector. *TEKNIKA Volume 6 Nomor 1*.
- Saurkar, A. V., Pathare, K. G., & Gode, S. A. (2018). An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering Volume 4 Issue 4*.
- Sembodo, J. E., Setiawan, E. B., & Baizal, Z. A. (2016). Data Crawling Otomatis pada Twitter. *Indonesia Symposium on Computing (IndoSC 2016)*.
- Singrodia, V., Mitra, A., & Paul, S. (2019). A Review on Web Scrapping and its Applications. *International Conference on Computer Communication and Informatics (ICCCI-2019)*.
- Sukendar, M. U. (2016). Pola Komunikasi Word Of Mouth (WoM) Brand Indosat Ooredoo Melalui Media Twitter. *Jurnal IKON Prodi D3 Komunikasi Massa–Politeknik Indonusa Surakarta Vol. 2 No. 4*.
- Twitter. (tanpa tahun). *Counting characters*. Diambil kembali dari Developer Twitter: <https://developer.twitter.com/en/docs/basics/counting-characters>