

**PENERAPAN TEKNIK *WEB SCRAPING* & METODE *TERM FREQUENCY - INVERS DOCUMENT FREQUENCY (TF-IDF)*
PADA PENCARIAN LOWONGAN KERJA**

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh:

Rusmansyah Putra NH
NIM : 09021381722122

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN SKRIPSI

**PENERAPAN TEKNIK *WEB SCRAPING* DAN METODE *TERM FREQUENCY – INVERS DOCUMENT FREQUENCY (TF-IDF)* PADA
PENCARIAN LOWONGAN KERJA**

Oleh:

Rusmansyah Putra NH

NIM: 09021381722122

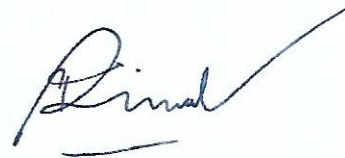
Palembang, Januari 2022

Pembimbing I



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

Pembimbing II



Mastura Diana Marieska, M.T
NIP. 198603212018032001

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Kamis tanggal 06 Januari 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rusmansyah Putra NH
NIM : 09021381722108
Judul : Penerapan Teknik *Web Scraping* dan Metode *Term Frequency – Invers Document Frequency (TF-IDF)* pada Pencarian Lowongan Kerja

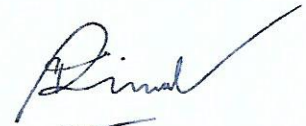
1. Pembimbing I

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



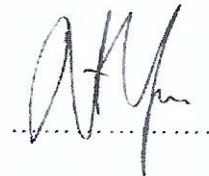
2. Pembimbing II

Mastura Diana Marieska, M.T
NIP. 198603212018032001



3. Penguji I

Novi Yusliani, M.T
NIP. 198211082012122001

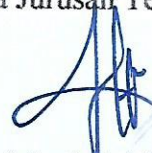


4. Penguji II

Desty Rodiah, M.T
NIP. 198912212020122011



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Rusmansyah Putra NH
NIM : 09021381722122
Judul : Penerapan Teknik *Web Scraping* dan Metode *Term Frequency – Invers Document Frequency* (TF-IDF) pada Pencarian Lowongan Kerja
Hasil Pengecekan Software iThenticate/Turnitin : 12%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya akan bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Januari 2022



Rusmansyah Putra NH
NIM. 09021381722122

MOTTO DAN PERSEMBAHAN

MOTTO:

تفاعل عندما تصعب عليك الأمور ، فإن الله تعالى أقسم مرتين فإنّ مع العُسْر يُسْرًا ، إنّ مع العسر يسرا

“Optimislah saat segala urusan terasa sulit bagimu, karena Allah telah bersumpah dua kali `Sesungguhnya bersama kesulitan ada kemudahan, sesungguhnya bersama kesulitan ada kemudahan`”

“Tetap Semangat”

Kupersembahkan Karya Tulis ini kepada:

- **Allah SWT.**
- **Orang tua dan saudara kandung ku**
- **Keluarga besar ku**
- **Dosen pembimbing**
- **Sahabat dan teman seperjuangan ku**
- **Fakultas Ilmu Komputer,
Universitas Sriwijaya**

ABSTRACT

According to data from the Badan Pusat Statistik (BPS, 2020) there has been an increase in the number of open unemployment coming from graduates of all levels of education. This study aims to simplify the job search process by combining and recommending job vacancies information from several sites using Web Scraping techniques and the Term Frequency–Inverse Document Frequency (TF-IDF) method. The concept of the Web Scraping technique is an automated data mining agent on web pages via DOM-HTML. Assisted by the TF-IDF method which evaluates the relationship of words in a document with keywords by assigning a weighted value to each word. Based on the results of testing using Usability Testing which produces an average value of 4.254 and included in the category very good overall so that this system is considered useful for users.

Keywords: Job Vacancies Information, Web Scraping, Dom-Html, Term Frequency–Inverse Document Frequency (TF-IDF), Usability Testing.

ABSTRAK

Menurut data Badan Pusat Statistik (BPS, 2020) terjadi peningkatan jumlah pengangguran terbuka yang berasal dari lulusan semua jenjang pendidikan. Penelitian ini bertujuan untuk mempermudah proses pencarian pekerjaan dengan menggabungkan dan merekomendasikan informasi lowongan kerja dari beberapa situs menggunakan teknik *Web Scraping* dan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Konsep dari teknik *Web Scraping* adalah agen penambang data otomatis pada halaman situs melalui DOM-HTML. Dibantu dengan metode TF-IDF yang mengevaluasi keterkaitan kata dalam dokumen dengan kata kunci dengan memberikan bobot nilai di setiap kata. Berdasarkan hasil pengujian menggunakan *Usability Testing* yang menghasilkan nilai rata-rata sebesar 4,254 dan masuk kedalam kategori sangat baik secara keseluruhan sehingga sistem ini dinilai berguna bagi pengguna.

Kata kunci: Informasi Lowongan Kerja, *Web Scraping*, Dom-Html, *Term Frequency–Inverse Document Frequency* (Tf-Idf), *Usability Testing*.

KATA PENGANTAR

Bismillahirrahmamanirrahim. Alhamdulillahirrabbi'l'amin. Puji dan syukur penulis panjatkan kepada Allah SWT. Atas segala rahmat, nikmat, dan karunia-Nya lah penulis dapat menyelesaikan penyusunan skripsi ini dengan judul **“Penerapan Teknik *Web Scraping* dan Metode *Term Frequency – Invers Document Frequency (TF-IDF)* pada Pencarian Lowongan Kerja”**. Shalawat diiringi salam tak lupa penulis hadiahkan kepada baginda Nabi Muhammad SAW. Semoga penulis, ibu dan ayah penulis, kakek dan nenek penulis, saudara penulis, sahabat dan teman penulis, guru, dosen dan kaum muslimin & muslimat mendapat syafaat Beliau di Yaumul Mahsyar kelak. Aamiin ya Rabbal'Alamin. Skripsi ini disusun dan diajukan untuk memenuhi syarat perolehan gelar sarjana (S.Kom) pada Fakultas Ilmu Komputer di Universitas Sriwijaya.

Untuk selanjutnya penulis mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu :

1. Orang tuaku; Naiman dan Rusnah, yang sangat berperan dalam membantu penulis, dengan mendo'akan dan memberi dukungan baik dalam segi moril maupun materi kepada penulis.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M.Kom, selaku Ketua Jurusan Teknik Informatika sekaligus Dosen Pembimbing I penulis, yang telah sangat banyak membantu, membimbing, mengarahkan, memberi masukan dan saran dalam penyelesaian skripsi ini.
4. Ibu Mastura Diana Marieska, M.T selaku Dosen Pembimbing II penulis, yang telah sangat banyak membantu dalam membimbing, mengarahkan, membantu, memberi masukan dan saran dalam penyelesaian skripsi ini.
5. Ibu Desty Rodiah, M.T, selaku Dosen Penguji sekaligus Dosen Pembimbing Akademik penulis, yang telah memberi saran, masukan dan bimbingannya selama masa perkuliahan sampai penulisan skripsi. Semoga Allah SWT meridhoi dan membalas kebaikannya.
6. Ibu Novi Yusliani, M.T, selaku Dosen Penguji penulis, yang telah memberi saran dan masukan agar penulisan skripsi ini dapat menjadi lebih baik. Semoga Allah SWT meridhoi dan membalas kebaikannya.
7. Seluruh Dosen yang telah mengajarkan penulis dengan memberikan informasi, pembelajaran, ilmu berharga dalam dunia perkuliahan penulis.
8. Mbak Wiwin Juliani, selaku Admin jurusan Teknik Informatika Bilingual, yang membantu dalam proses administrasi selama perkuliahan penulis.
9. Keluarga Besarku yang telah memberikan dukungan, hiburan, arahan, bahkan materi yang membuat penulis sangat terbantu dalam menyelesaikan perkuliahan dan skripsi ini.
10. M. Aldi Ariqi dan Ahmad Emir Alfatah, selaku teman sekelas penulis di jurusan Teknik Informatika ini yang sangat baik mau meluangkan waktunya untuk membantu dalam proses penyelesaian perangkat lunak untuk skripsi penulis. Semoga Allah SWT meridhoi dan membalas kebaikan kalian.

11. Bunga Ayu Ferdiyanti, selaku teman sekelas penulis di jurusan Teknik Informatika ini yang sangat baik mau membantu dalam perihal pengurusan berkas-berkas maupun materil yang dibutuhkan. Semoga Allah SWT meridhoi dan membalas kebaikanmu.
12. Robi Hidayat, selaku teman begadang untuk kepentingan menyelesaikan pengerjaan skripsi penulis. Semoga Allah SWT meridhoi dan membalas kebaikanmu.
13. M. Imam Renaldy Gumay, Zhafirah Rahmadini, Stefany Naomi B.S, Sausan Syahirah, Rezki Adina, Wiwik Anum Puspita teman satu kelas penulis di jurusan Teknik Informatika yang menjadi teman dekat selama perkuliahan dan teman “Skripbar” yang memunculkan banyak motivasi dalam pengerjaan skripsi ini. Semoga Tuhan meridhoi dan membalas kebaikan kalian.
14. Yasmin Azzahra Lubis, selaku teman seangkatan penulis di jurusan Teknik Informatika yang telah memberikan ide langkah untuk memulai skripsi penulis dan juga sangat membantu dalam proses pembangunan script python. Semoga Allah SWT meridhoi dan membalas kebaikanmu.
15. Nanda Tirana Mahesa dan Dewi Chayanti, selaku teman seangkatan penulis di jurusan Teknik Informatika yang telah memperkenalkan dan membantu pengerjaan script python. Semoga Allah SWT meridhoi dan membalas kebaikan kalian.
16. Mega Pertiwi, selaku teman seangkatan penulis di jurusan Teknik Informatika sekaligus “Peri Sihir” dan partner “Food Hunting” yang telah meluangkan waktunya untuk mendengarkan keluh-kesah dan menemani healing. Semoga Allah SWT meridhoi dan membalas kebaikan kalian.
17. Teman – teman satu kelas “Tibil Baper” dan satu jurusan Teknik Informatika. Selalu kompak dan sukses selalu.
18. Dinas Pengembangan Sumber Daya Mahasiswa HMIF FASILKOM 2019 yang tetap selalu saling mendukung satu sama lain.
19. Keluarga besar BEM KM FASILKOM Kabinet Samudra 2018 dan Kelurga besar HMIF 2018. Semoga sukses dan Sehat selalu.
20. Last but not least, saya ingin mengucapkan terima kasih. Saya ingin berterima kasih kepada saya karena percaya pada saya. Saya ingin berterima kasih kepada saya karena telah melakukan semua kerja keras ini. Saya ingin berterima kasih kepada saya karena tidak memiliki hari libur. Saya ingin berterima kasih untuk, karena tidak pernah berhenti. Saya ingin berterima kasih kepada saya karena selalu menjadi pemberi. Dan mencoba memberi lebih dari yang saya terima. Saya ingin berterima kasih kepada saya karena mencoba melakukan lebih banyak benar daripada salah. Saya ingin berterima kasih kepada saya karena telah menjadi saya setiap saat. Semoga Allah SWT meridhoi jalanku.

21. Beserta semua & seluruh pihak yang membantu yang tidak dapat penulis tulis satu persatu dan yang mungkin belum penulis tuliskan mohon dimaafkan.

Penulis menyadari bahwa masih banyak kekurangan dan kesalahan dalam penyusunan Skripsi ini, dan Skripsi ini jauh dari kata sempurna karena keterbatasan ilmu, kemampuan, dan pengalaman yang dimiliki. Penulis berharap semoga Skripsi ini dapat membantu dan bermanfaat bagi semua pihak yang memerlukan.

Palembang. Januari 2022

Rusmansyah Putra NH

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah	I-6
1.7 Sistematika penulisan	I-7
1.8 Kesimpulan	I-8
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 <i>Document Object Model-Hyperteks Markup Language (DOM-HTML)</i>	II-1
2.2.2 <i>Web Scraping</i>	II-2
2.2.3 <i>Term Frequency – Invers Document Frequency (TF-IDF)</i> ..	II-8
2.2.4 <i>Rational Unified Process (RUP)</i>	II-14
2.2.5 <i>Usability Testing</i>	II-14
2.3 Penelitian Lain yang Relevan	II-17
2.3.1 Penerapan <i>Web Scraping</i> Pada Situs <i>Company Profile</i>	II-14
2.3.2 Rekomendasi Pekerjaan dari Portal Bursa Lowongan Kerja Memanfaatkan Fitur <i>Cosine Similarity</i> dan <i>Simple Additive Weighting</i>	II-14
2.3.3 Klasifikasi Berita Online dengan Menggunakan Pembobotan TF-IDF dan <i>Cosine Similarity</i>	II-14
2.4 Kesimpulan	II-20

BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.2.1 Jenis Data	III-1
3.2.2 Sumber Data	III-2
3.2.3 Metode Pengumpulan Data	III-3
3.3 Tahapan Penelitian	III-3
3.3.1 Kriteria Pengujian	III-11
3.3.1.1 Kasus Uji	III-12
3.3.1.2 Usability Testing	III-13
3.3.2 Format Data Pengujian	III-14
3.3.3 Alat yang Digunakan	III-18
3.3.4 Pengujian Penelitian	III-18
3.3.5 Analisis Hasil Pengujian dan Membuat Kesimpulan	III-20
3.4 Metode Pengembangan Perangkat Lunak.....	III-21
3.4.1 Fase Insepsi	III-21
3.4.2 Fase Elaborasi.....	III-22
3.4.3 Fase Konstruksi	III-22
3.4.4 Fase Transisis	III-23
3.5 Manajemen Proyek Penelitian	III-23
3.6 Kesimpulan	III-28
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....	IV-1
4.1 Pendahuluan	IV-1
4.2 <i>Rational Unified Process</i> (RUP).....	IV-1
4.2.1 Analisa Kebutuhan Perangkat Lunak.....	IV-1
4.2.2 Perancangan Perangkat Lunak	IV-15
4.2.2.1 <i>Use Case</i>	IV-15
4.2.2.2 <i>Analysis Class Diagram</i>	IV-24
4.2.2.3 <i>Diagram Activity</i>	IV-25
4.2.2.4 <i>Sequence Diagram</i>	IV-30
4.2.2.5 <i>Class Diagram</i>	IV-32
4.2.2.6 Desain Antarmuka.....	IV-34
4.2.3 Implementasi Perangkat Lunak.....	IV-37
4.2.3.1 Implementasi Kelas	IV-37
4.2.3.2 Implementasi Antarmuka	IV-39
4.2.4 Pengujian Perangkat Lunak.....	IV-42
4.2.4.1 Rencana Pengujian	IV-42
4.2.4.2 Kasus Uji.....	IV-46
4.3 Kesimpulan	IV-57
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan	V-1

5.2 Data Hasil Penelitian/Percobaan	V-1
5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Data Hasil Pengujian.....	V-5
5.2.3 Analisis Hasil Penelitian	V-16
5.3 Kesimpulan	V-21
BAB VI KESIMPULAN DAN SARAN.....	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-2
DAFTAR PUSTAKA.....	xviii

DAFTAR TABEL

	Halaman
Tabel II - 1. Hasil Hitung Nilai TF	II-11
Tabel II - 2. Hasil Hitung Nilai DF dan IDF	II-12
Tabel II - 3. Hasil Hitung Nilai TF-IDF	II-13
Tabel II - 4. Rentang Skala Nilai Rata-rata	II-17
Tabel III - 1. Rancangan Tabel Basis data Hasil Ekstraksi	III-15
Tabel III - 2. Rancangan Tabel Hasil Perhitungan Manual.....	III-16
Tabel III - 3. Rancangan Tabel Skenario Uji	III-16
Tabel III - 4. Rancangan Tabel Kuisisioner	III-17
Tabel III - 5. Rancangan Tabel Perbandingan Hasil Perhitungan Manual dan Hasil Perhitungan Perangkat Lunak	III-19
Tabel III - 6. Tabel <i>Work Breakdown Structure</i> (WBS) dari Manajemen Proyek Penelitian.....	III-24
Tabel IV - 1. Sampel Data Hasil Implementasi <i>Web Scraping</i>	IV-3
Tabel IV - 2. Data <i>Stamming</i> dari Sampel Data.....	IV-6
Tabel IV - 3. Hasil Hitung Nilai TF dari Sampel Data.....	IV-11
Tabel IV - 4. Hasil Hitung Nilai TF-IDF dari Sampel Data	IV-12
Tabel IV - 5. Hasil Total Hitung TF-IDF dari Sampel Data.....	IV-13
Tabel IV - 6. Definisi Aktor.....	IV-16
Tabel IV - 7. Definisi <i>Use Case</i>	IV-17
Tabel IV - 8. Skenario <i>Use Case</i> Melakukan Pencarian dengan Memasukkan Kata Kunci.....	IV-19
Tabel IV - 9. Skenario <i>Use Case</i> Mengunjungi Langsung ke Situs Lowongan Kerja	IV-20
Tabel IV - 10. Update Status Informasi Lowongan Kerja.....	IV-21
Tabel IV - 11. Skenario <i>Use Case</i> Scraping Situs Lowongan Kerja	IV-23
Tabel IV - 12. Daftar Implementasi Kelas.....	IV-37

Tabel IV - 13. Skenario Pengujian Melakukan Pencarian dengan Memasukkan Kata Kunci	IV-43
Tabel IV - 14. Skenario Pengujian Mengunjungi Langsung ke Situs Lowongan Kerja	IV-43
Tabel IV - 15. Skenario Pengujian <i>Update</i> Data Informasi Lowongan Kerja	IV-44
Tabel IV - 16. Skenario Pengujian <i>Scraping</i> Situs Lowongan Kerja	IV-44
Tabel IV - 17. Pengujian <i>use case</i> Melakukan Pencarian dengan Memasukkan Kata Kunci	IV-46
Tabel IV - 18. Tabel Perbandingan Hasil Perhitungan Manual dan Hasil Perhitungan Perangkat Lunak.....	IV-47
Tabel IV - 19. Pengujian <i>use case</i> Mengunjungi Langsung ke Situs Lowongan Kerja	IV-52
Tabel IV - 20. Pengujian <i>use case Update</i> Data Informasi Lowongan Kerja.	IV-53
Tabel IV - 21. Pengujian <i>use case Scraping</i> Situs Lowongan Kerja	IV-55
Tabel V - 1. Task Skenario untuk <i>Usability Testing</i>	V-2
Tabel V - 2. Pertanyaan dan Pengkodean Pada Kuisisioner	V-3
Tabel V - 3. Tabel Data Persentase Hasil Task Per-Responden	V-17
Tabel V - 4. Hasil Rekap Nilai <i>Usability Testing</i>	V-19

DAFTAR GAMBAR

	Halaman
Gambar II - 1. Halaman Awal Situs Jobstreet (www.jobstreet.co.id)	II-6
Gambar II - 2. Halaman Awal Situs Glints (glints.com)	II-7
Gambar II - 3. Halaman Awal Situs Glints (glints.com)	II-7
Gambar II - 4. Halaman Awal Situs Karir.com	II-8
Gambar II - 5. Arsitektur <i>Rational Unified Process</i> (RUP) (Manalil, 2010)	II-15
Gambar III - 1. Diagram Kerangka Kerja Penelitian.....	III-4
Gambar III - 2. Diagram Detil Analisis Kebutuhan	III-5
Gambar III - 3. Diagram Detil Pengembangan Perangkat Lunak	III-7
Gambar III - 4. Diagram Detil Pengujian	III-11
Gambar IV - 1. Diagram <i>Use case</i>	IV-15
Gambar IV - 2. <i>Analysis Class Diagram</i> Melakukan Pencarian dengan Memasukkan Kata Kunci	IV-24
Gambar IV - 3. <i>Analysis Class Diagram</i> Mengunjungi Langsung ke Situs Lowongan Kerja	IV-25
Gambar IV - 4. <i>Analysis Class Diagram Update</i> Informasi Lowongan Kerja dan <i>Scraping</i> Situs Lowongan Kerja.....	IV-25
Gambar IV - 5. Diagram <i>Activity</i> Melakukan Pencarian dengan Memasukkan Kata Kunci	IV-26
Gambar IV - 6. Diagram <i>Activity</i> Mengunjungi Langsung ke Situs Lowongan Kerja	IV-27
Gambar IV - 7. Diagram <i>Activity Update</i> Status Data Informasi Lowongan Kerja..	IV-28
Gambar IV - 8. Diagram <i>Activity Scraping</i> Situs Lowongan Kerja.....	IV-29
Gambar IV - 9. <i>Sequence Diagram</i> Melakukan Pencarian dengan Memasukkan Kata Kunci	IV-30

Gambar IV - 10. <i>Sequence Diagram</i> Mengunjungi Langsung ke Situs Lowongan Kerja	IV-31
Gambar IV - 11. <i>Sequence Diagram</i> Update Data Informasi Lowongan Kerja.....	IV-31
Gambar IV - 12. <i>Sequence Diagram</i> Scraping Situs Lowongan Kerja ..	IV-32
Gambar IV - 13. <i>Class Diagram</i> TF-IDF	IV-33
Gambar IV - 14. <i>Class Diagram</i> Scraping dan Update	IV-33
Gambar IV - 15. Rancangan Desain Halaman Awal Situs	IV-34
Gambar IV - 16. Rancangan Desain Halaman Penampil List Rekomendasi Berdasarkan TF-IDF.....	IV-35
Gambar IV - 17. Rancangan Desain Halaman <i>About</i>	IV-36
Gambar IV - 18. Rancangan Desain Halaman <i>Help</i>	IV-36
Gambar IV - 19. Halaman Awal Situs	IV-40
Gambar IV - 20. Halaman Penampil List Rekomendasi Berdasarkan TF-IDF	IV-40
Gambar IV - 21. Halaman Penampil List Perhitungan Pembobotan Berdasarkan TF-IDF.....	IV-41
Gambar IV - 22. Halaman <i>About</i>	IV-41
Gambar IV - 23. Halaman <i>Help</i>	IV-42

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan akan membahas tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian serta batasan masalah. Penjelasan secara umum mengenai penelitian akan dibahas pada bab ini.

Penjelasan akan dimulai dengan membahas latar belakang mengenai penelitian tentang Implementasi Teknik *Web Scraping* pada Pencarian Lowongan Kerja Menggunakan Metode *Term Frequency Invers Document Frequency* (TF-IDF).

1.2 Latar Belakang

Ketenagakerjaan adalah salah satu sektor vital yang berdampak pada perekonomian dan kemakmuran suatu negara. Menurut data Badan Pusat Statistik yang bersumber dari Survei Angkatan Kerja Nasional bahwa jumlah angkatan kerja meningkat sebanyak 1.62 juta pada february 2020 (Sakernas, 2020). Namun sayangnya Indonesia mengalami beberapa masalah dikarenakan terjadi peningkatan pengangguran terbuka, contohnya untuk bulan february 2020 sebanyak 6.92 juta pengangguran terbuka berasal dari lulusan semua jenjang pendidikan yang secara khusus sebanyak 170.531 berasal dari Sumatera Selatan.

Proses pencarian pekerjaan biasanya dilakukan secara *online* oleh pencari kerja karena proses yang lebih mudah dibandingkan dengan media konvensional.

Menurut Bart sebanyak 55 persen pencari kerja menggunakan situs umum sebagai media mencari pekerjaan mereka. Data ini belum ditambah dengan penggunaan situs kerja *online* sebesar 58 persen (Turczynski, 2021). Namun sayangnya untuk saat ini pencari masih harus membuka situs informasi lowongan pekerjaan satu persatu yang berarti membutuhkan waktu tambahan untuk proses mendapatkan pekerjaan. Penelitian ini bertujuan untuk mempermudah proses pencarian pekerjaan karena akan menggabungkan informasi lowongan kerja dari beberapa situs menggunakan teknik *Web Scraping*.

Web Scraping adalah teknik untuk mendapatkan informasi dari halaman web yang umumnya dibangun menggunakan bahasa markup seperti DOM-HTML. *Web Scraping* adalah agen penambang data otomatis pada suatu situs yang sesedikit mungkin melibatkan manusia dalam prosesnya. Mungkin teknik *scraping* sama tuanya dengan internet itu sendiri. Praktik yang lebih dulu dikenal sebelum *web scraping* adalah *screen scraping*, yaitu praktik mengekstrak data dari representasi visual yang terjadi pada masa awal komputasi sekitar tahun 1960-1980an dan biasanya bermuara pada "terminal" sederhana berbasis teks (vanden Broucke & Baesens, 2018).

Sedangkan DOM adalah *Document Object Model*, yaitu *application programming interface* (API) atau model penulisan standar yang mendefinisikan dokumen XML atau HTML pembentuk situs secara rinci. Istilah "*Document*" dalam DOM mengandung arti sebagai data bukan sebagai dokumen. Standarisasi DOM-HTML mengartikan bahwa apa pun yang ditemukan dalam dokumen HTML dapat diakses, diubah, dihapus, atau ditambahkan menggunakan DOM

(W3C, 2000). Sedangkan HTML atau *HyperText Markup Language* adalah blok pembangun situs yang paling dasar yang menggunakan "markup" sebagai penampil dalam peramban untuk anotasi teks, gambar, dan konten lainnya yang ditandai dengan "element" khusus. Elemen ini dibedakan dengan "tags" dari teks lain yang terdiri dari nama elemen serta dikelilingi oleh tanda "<" dan ">" (MDN, 2021). Markup "tags" inilah yang akan dijadikan sebagai acuan data yang akan diambil dalam teknik *web scraping* dalam penelitian ini.

Dibantu dengan metode TF-IDF yang mengevaluasi seberapa relevan suatu kata dengan kata kunci. Dalam buku berjudul "*Encyclopedia of Machine Learning and Data Mining*" yang dipublikasikan oleh *Springer US* (Hodson, 2017) dijelaskan bahwa TF-IDF adalah skema pembobotan istilah sebagai klasifikasi, pengelompokan, visualisasi, pengambilan, dan lain-lain. Metode TF-IDF biasa digunakan untuk menentukan keterhubungan kata (term) sebagai bentuk pengelompokan dengan memberikan bobot nilai di setiap kata. Metode ini adalah hasil menggabungkan dua konsep, yaitu: *Term Frequency* yang menghitung frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan *Invers Document Frequency* yang meng-inversi frekuensi dokumen yang mengandung kata tersebut (Herwijayanti et al., 2018).

Teknik *Web Scraping* ini telah diimplementasikan pada sejumlah penelitian, pun juga untuk metode TF-IDF telah dikembangkan oleh sejumlah peneliti. Sebagai contoh Dermawan dkk mengekstrak data hirarki *Kingdom Animalia* berdasarkan pola navigasi dan struktur HTML dari situs a-z-animal.com untuk mendapatkan data yang menunjukkan bahwa *Web Scraping* dapat menangani proses ekstraksi

konten situs a-z-animal.com berupa data takson, deskripsi, dan gambar dari berbagai macam hewan (Darmawan et al., 2017). Dan pada penelitian Herwijayanti dkk tentang klasifikasi berita daring dengan perangkat lunak metode TF-IDF dan *Cosine Similarity* yang memberi kesimpulan tingkat akurasi sebesar 91.25 persen sampai 100 persen pada kategori data hasil sistem sesuai dengan data asli (Herwijayanti et al., 2018).

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana cara mengimplementasikan teknik *Web Scraping* pada situs penyedia informasi lowongan pekerjaan serta implementasi metode TF-IDF? Menjawab rumusan masalah tersebut, dibawah ini akan diuraikan beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana implementasi *Web Scraping* untuk mendapatkan dataset informasi lowongan pekerjaan dari masing-masing situs lowongan kerja yang tidak menyediakan layanan berbagi data?
2. Bagaimana metode TF-IDF dapat terimplementasi untuk pembobotan dalam memberikan rekomendasi lowongan pekerjaan?
3. Bagaimana cara penyajian informasi lowongan kerja hasil dari teknik *Web Scraping* dan rekomendasi menggunakan pembobotan metode TF-IDF?

1.4 Tujuan Penelitian

Penelitian ini diharapkan untuk menyelesaikan rumusan masalah yang ditemukan, yaitu terimplementasinya teknik *Web Scaping* pada situs penyedia informasi lowongan pekerjaan serta implementasi metode TF-IDF untuk pembobotan rekomendasi berupa:

1. Menghasilkan sistem rekomendasi lowongan pekerjaan hasil implementasi metode TF-IDF dalam pembobotan berdasarkan kata kunci yang dimasukkan.
2. Mendapatkan hasil penyajian lowongan kerja dari dataset hasil teknik *web scraping* yang ditampilkan sesuai rekomendasi menggunakan pembobotan metode TF-IDF.

1.5 Manfaat Penelitian

Penyelesaian rumusan masalah pada penelitian ini diharapkan untuk mampu memberi manfaat berupa:

1. Mempermudah para pencari kerja untuk memilih lowongan pekerjaan berdasarkan kata kunci yang dimasukkan.
2. Memberikan rekomendasi lowongan pekerjaan yang sesuai dengan kata kunci yang dimasukkan.
3. Menambah opsi rujukan untuk penelitian lain yang sejenis.

1.6 Batasan Masalah

Penelitian ini memiliki batasan masalah berupa:

1. Perangkat lunak hanya ditujukan sebagai media rekomendasi lowongan pekerjaan.
2. Implementasi teknik *Web Scraping* dan metode TF-IDF dilakukan secara terpisah, yakni:
 - a) Teknik *Web Scraping* berupa sistem perangkat lunak yang akan dijalankan secara otomatis menggunakan *Task Scheduler* atau *Cron Job*.
 - b) Metode TF-IDF diimplementasikan pada perangkat lunak situs utama.
3. *Scraping* hanya menggunakan kata kunci yang telah ditentukan pada basis data dan akan diperbaharui secara otomatis sesuai dengan jadwal yang telah ditetapkan.
4. *Scraping* hanya dilakukan pada situs Jobstreet (www.jobstreet.co.id), Glints (glints.com), dan Karir (www.karir.com).
5. Perangkat lunak dibangun berdasarkan anggapan mampu mengekstrak konten situs berupa judul lowongan pekerjaan, nama perusahaan pemberi lowongan kerja, lokasi perusahaan, detail informasi lowongan pekerjaan, dan link informasi lowongan pekerjaan.
6. Fungsionalitas perangkat lunak dirancang sebagai berikut:
 - a) Perangkat lunak mampu mengekstrak konten lowongan kerja.

- b) Perangkat lunak dapat mengurutkan rekomendasi pekerjaan berdasarkan kata kunci yang dimasukkan.
- c) Perangkat lunak dapat membawa pengguna ke halaman situs lowongan kerja yang dipilih untuk melanjutkan proses pendaftaran lowongan kerja.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I. PENDAHULUAN:

Bab ini berisi penjabaran mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II . KAJIAN LITERATUR

Pembahasan mengenai dasar-dasar teori yang digunakan, seperti definisi-definisi *Web Scraping*, metode *Term Frequency Invers Document Frequency* (TF-IDF), software, serta beberapa kajian literatur mengenai penelitian lain yang relevan pada penelitian ini akan dibahas pada bab ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pembahasan mengenai perancangan dan lingkungan implementasi teknik pengambilan data, implementasi program rekomendasi, hasil perancangan, hasil eksekusi dan pengujian hasil implementasi.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini akan berisi penyajian hasil dari pengujian implementasi berdasarkan langkah-langkah pada perencanaan pengujian. Hasil akan dianalisis sebagai basis dari kesimpulan yang akan diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan berisikan kesimpulan dari semua uraian pada bab-bab sebelumnya serta berisi saran-saran yang diharapkan dapat berguna dalam penelitian selanjutnya mengenai implementasi teknik *web scraping* dan metode TF-IDF khususnya pada pencarian lowongan kerja.

1.8 Kesimpulan

Dapat diambil simpulan untuk bab ini bahwa masalah yang harus diselesaikan melalui penelitian ini adalah bagaimana membangun sebuah perangkat lunak rekomendasi lowongan kerja dengan menerapkan Teknik *Web Scraping* dan Metode *Term Frequency - Invers Document Frequency* (TF-IDF) yang dapat membantu pengguna dalam memilah Pencarian Lowongan Kerja.

DAFTAR PUSTAKA

- B, M. R., Pratama, Y., & Fachruddin. (2017). Penerapan Web Scraping Pada Website Company Profile. *Sekolah Tinggi Ilmu Komputer Dinamika Bangsa*, 4(4), 37–43.
- Darmawan, I. P. A., Piarsa, I. N., & Dharmadi, I. P. A. (2017). Ekstrak Hirarki Data Dari Situs Web A-Z Animals Menggunakan Web Scraping. *LONTAR KOMPUTER*, 8(3), 166–177.
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 306–312.
- Hodson, J. (2017). *Encyclopedia of Machine Learning and Data Mining* (C. Sammut & G. I. Webb (eds.)). Springer US. <https://doi.org/10.1007/978-1-4899-7687-1>
- Huda, N. (2019). Implementasi Metode Usability Testing Dengan System Usability Scale Dalam Penilaian Website Rs Siloam Palembang. *Klik - Kumpulan Jurnal Ilmu Komputer*, 6(1), 36. <https://doi.org/10.20527/klik.v6i1.177>
- Manalil, J. (2010). RATIONAL UNIFIED PROCESS. *Computer*, August.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). Scoring, Term Weighting and the Vector Space Model. In *Introduction to Information Retrieval* (1st ed., Issue c, pp. 120–126). Cambridge University Press. <http://www-nlp.stanford.edu/IR-book/>
- Manning, C. D., Raghavan, P., & Schütze, H. (2018). *TF-IDF Weighting*. Cambridge University Press. <https://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html>
- MDN. (2021). *HTML Tutorial Pemula*. MDN Web Docs. <https://developer.mozilla.org/id/docs/Web/HTML>
- Ningsih, R., Azhar, Y., & Munarko, Y. (2020). Rekomendasi Lowongan Pekerjaan dari Portal Bursa Lowongan Kerja Memanfaatkan Cosine Similarity dan Simple Additive Weighting. *Repositor*, 2(5), 601. <https://doi.org/10.22219/repositor.v2i5.58>
- Nurhadryani, Y., Sianturi, S. K., Hermadi, I., & Khotimah, H. (2013). Pengujian Usability untuk Meningkatkan Antarmuka Aplikasi Mobile. *Jurnal Ilmu Komputer Dan Agri-Informatika*, 2(2), 83. <https://doi.org/10.29244/jika.2.2.83-93>
- Rifqatusa'adah, F. L. (2017). *EVALUASI USABILITY BERDASARKAN ISO/IEC 9126 DAN NIELSEN MODEL MENGGUNAKAN METODE USABILITY*

TESTING (STUDI KASUS: APLIKASI MOBILE REBLOOD).

- Sakernas. (2020). Tingkat Pengangguran Terbuka Berdasarkan Tingkat Pendidikan. In *Bps.Go.Id*. <https://bps.go.id/indicator/6/1179/1/Tingkat-Pengangguran-Terbuka-Berdasarkan-Tingkat-Pendidikan.html>
- Sugiharto, T., Kadir, A., & Ferdiana, R. (2014). Perancangan Mobile Application Untuk Penanganan Tindak Pidana Pelanggaran Lalu-Lintas Menggunakan Metode Rational Unified Process. *Seminar Nasional Teknologi Informasi Dan Multimedia*, 13–20.
- Sukmasetya, P., Setiawan, A., & Arumi, E. R. (2020). Penggunaan Usability Testing Sebagai Metode Evaluasi Website Krs Online Pada Perguruan Tinggi. *JST (Jurnal Sains Dan Teknologi)*, 9(1), 58–67. <https://doi.org/10.23887/jst-undiksha.v9i1.24691>
- Turczynski, B. (2021). *2021 HR Statistics: Job Search, Hiring, Recruiting & Interviews*. Zety. <https://zety.com/blog/hr-statistics>
- vanden Broucke, S., & Baesens, B. (2018). *Practical Web Scraping for Data Science* (1st ed.). Apress. <https://doi.org/https://doi.org/10.1007/978-1-4842-3582-9>
- W3C. (2000). *Document Object Model (DOM) Level 2 Core Specification*. World Wide Web Consortium (W3C). <https://www.w3.org/TR/DOM-Level-2-Core/Overview.html#contents>
- Wignjosoebroto, S., Sudiarno, A., & Harenda, D. (2011). *PERANCANGAN INTERFACE PROTOTYPE WEB BERDASARKAN PADA ASPEK USABILITY (Studi Kasus : Laboratorium Ergonomi dan Perancangan Sistem Kerja Teknik Industri ITS)*. 1–14. <http://digilib.its.ac.id/public/ITS-Undergraduate-8877-2504100030-Paper.pdf>