

SKRIPSI
SISTEM EKSTRAKSI DAFTAR RIWAYAT HIDUP
MENGGUNAKAN *NAMED ENTITY RECOGNITION*



RISKY ELIANA DEWI

NIM : 195611024

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2023

SKRIPSI
SISTEM EKSTRAKSI DAFTAR RIWAYAT HIDUP
MENGGUNAKAN *NAMED ENTITY RECOGNITION*

Diajukan sebagai salah satu syarat untuk menyelesaikan studi

Program Sarjana
Program Studi Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta



Disusun Oleh
RISKY ELIANA DEWI
NIM : 195611024

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2023

DAFTAR ISI

SKRIPSI	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
KATA PENGANTAR	ix
INTISARI	xi
ABSTRACT	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 TINJAUAN PUSTAKA DAN DASAR TEORI	6
2.1 Tinjauan Pustaka	6
2.2 Dasar Teori	8
2.2.1 Bahasa Pemrograman <i>Python</i>	8
2.2.2 Ekstraksi Informasi	9
2.2.3 Evaluasi Skor	9
2.2.4 <i>Framework Flask Python</i>	10
2.2.5 <i>Named Entity Recognition (NER)</i>	11
2.2.6 <i>PostgreSQL</i>	11
2.2.7 REST API	11
2.2.8 <i>Rule Based</i>	12

2.2.9 <i>SpaCy</i>	12
BAB 3 METODE PENELITIAN	14
3.1 Bahan/Data	14
3.2 Peralatan	14
3.2.2 Perangkat Keras	14
3.2.2 Perangkat Lunak	14
3.3 Prosedur dan Pengumpulan Data	15
3.4 Analisis dan Rancangan Sistem	16
BAB 4 IMPLEMENTASI DAN PEMBAHASAN	20
4.1 Implementasi Sistem	20
4.2 Pengujian	32
4.3 Pembahasan	34
BAB 5 PENUTUP	36
5.1 Kesimpulan	36
5.2 Saran	36
DAFTAR PUSTAKA	37

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Penelitian	16
Gambar 4.1 Daftar Riwayat Hidup dari Internet	20
Gambar 4.2 Daftar Riwayat Hidup Sesuai Format	21
Gambar 4.3 Kode Program Ekstraksi Teks	22
Gambar 4.4 Kode Program Cleansing	22
Gambar 4.5 Pemberian Label	23
Gambar 4.6 Anotasi	23
Gambar 4.7 Kode Program Mengubah Format SpaCy	24
Gambar 4.8 Kode Program Melatih Model NER	24
Gambar 4.9 Kode Program Menyimpan Model NER	25
Gambar 4.10 Kode Program Rule Based	26
Gambar 4.11 Kode Program Evaluasi Model NER	27
Gambar 4.12 Parsing Anotasi	27
Gambar 4.13 Kode Program Prediksi Deteksi Entitas	28
Gambar 4.14 Kode Program Evaluasi untuk Entitas Pendidikan	29
Gambar 4.15 Kode Program Evaluasi untuk Entitas Jurusan	29
Gambar 4.16 Kode Program Evaluasi untuk Entitas Pengalaman	29
Gambar 4.17 Kode Program Evaluasi untuk Entitas Tahun	29
Gambar 4.18 Kode Program Evaluasi untuk Entitas Teknis	30
Gambar 4.19 Kode Program Evaluasi untuk Entitas Non Teknis	30
Gambar 4.20 Kode Program Evaluasi untuk Entitas Bahasa	30
Gambar 4.21 Kode Program Evaluasi untuk Entitas Organisasi	30
Gambar 4.22 Kode Program Evaluasi untuk Entitas Prestasi	31
Gambar 4.23 Kode Program Pembuatan REST API	32
Gambar 4.24 Kode Program Membaca PDF di dalam Database	32
Gambar 4.25 Kode Program Keluaran API	33

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	7
Tabel 3.1 Entitas	18
Tabel 4.1 Nilai Akurasi	33
Tabel 4.2 Nilai Akurasi Per Entitas	34
Tabel 4.3 Nilai Akurasi 2	334

KATA PENGANTAR

Segala puji dan syukur atas kehadiran Allah SWT yang telah memberikan Rahmat dan Karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Sistem Ekstraksi Daftar Riwayat Hidup Menggunakan *Named Entity Recognition* (NER)”. Skripsi ini disusun untuk memenuhi tugas akhir sebagai mahasiswa Program Studi Sistem Informasi Universitas Teknologi Digital Indonesia Yogyakarta.

Penulis menyadari skripsi ini tidak dapat terselesaikan tepat waktu tanpa adanya bimbingan, bantuan, dukungan, dan doa dari berbagai pihak. Dengan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Bapak Ir. Totok Suprawoto, M.M., M.T. selaku Rektor Universitas Teknologi Digital Indonesia Yogyakarta.
2. Ibu Pulut Suryati, S.Kom., M.Cs selaku Ketua Program Studi Sistem Informasi Universitas Teknologi Digital Indonesia Yogyakarta.
3. Bapak Edy Prayitno S.Kom., M.Eng. selaku dosen pembimbing skripsi yang telah banyak memberikan bimbingan, pengarahan, saran dan motivasi dalam mengerjakan skripsi ini hingga selesai.
4. Ibu Dini Fakta Sari, S.T., M.T., Ibu Deborah Kurniawati, S.Kom., M.Cs., dan Bapak Rikie Kartadie, S.T., M.Kom. selaku dosen penguji yang telah memberikan kritik dan saran dalam menyelesaikan skripsi ini.
5. Seluruh Dosen di Program Studi Sistem Informasi Universitas Teknologi Digital Indonesia Yogyakarta yang selama ini telah membagikan ilmunya kepada penulis.

6. Bapak dan Ibu selaku orang tua yang selalu mendoakan penulis selama menempuh pendidikan.
7. Teman sekelompok yang turut memberikan masukan, dukungan dan membantu menyelesaikan skripsi ini.
8. Serta semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan dukungan dan membantu menyelesaikan skripsi ini.

Dalam pembuatan skripsi ini, penulis juga menyadari skripsi ini masih banyak kekurangan, baik dari materi, penulisan maupun dari segi penyajian karena keterbatasan dan kemampuan penulis. Oleh karena itu, penulis sangat mengharapkan saran dan kritik untuk kesempurnaan skripsi ini. Semoga skripsi ini dapat memberikan manfaat maupun inspirasi bagi penulis dan bagi pembaca.

INTISARI

Kemajuan teknologi dan pertumbuhan ekonomi meningkatkan daya saing dalam bisnis. Perusahaan membutuhkan sumber daya manusia yang berkualitas dan kompeten untuk mengembangkan dan mencapai tujuan, karena kemampuan mereka secara signifikan mempengaruhi kesuksesan perusahaan. Rekrutmen adalah proses penting bagi perusahaan untuk mendapatkan tenaga kerja yang tepat untuk berbagai posisi, dengan tahap awal pelamar mengirimkan daftar riwayat hidup ke perusahaan. Daftar riwayat hidup adalah dokumen ringkas yang berisi kualifikasi, prestasi dan pengalaman pelamar yang relevan. Perusahaan perlu membacanya satu per satu untuk menemukan karyawan yang sesuai, namun proses ini bisa memakan waktu dan rawan terjadi kesalahan, sehingga berpotensi membatasi kesempatan untuk mendapatkan kandidat terbaik.

Penelitian ini bertujuan untuk mengembangkan sistem ekstraksi daftar riwayat hidup menggunakan *Named Entity Recognition* (NER). Penelitian ini membuat sistem untuk mengidentifikasi dan mengklasifikasikan isi daftar riwayat hidup secara otomatis menggunakan *Named Entity Recognition* (NER) dengan pendekatan *Rule Based*, sebuah komponen inti dari ekstraksi informasi yang memberi label pada data masukan. Penelitian ini melatih model untuk mendeteksi entitas dalam teks riwayat hidup, dengan hasil nilai *precision* 100%, *recall* 87% dan *f-score* 93%.

Kata Kunci: Daftar Riwayat Hidup, Ekstraksi Informasi, Rekrutmen, Named Entity Recognition (NER), Rule Based

ABSTRACT

Technological advances and economic growth increase competitiveness in business. Companies need qualified and competent human resources to develop and achieve goals, because their abilities greatly affect the success of the company. Recruitment is an important process for companies to get the right workforce for various positions, with the initial stage of applicants sending a curriculum vitae to the company. A curriculum vitae is a concise document containing an applicant's relevant qualifications, achievements and experience. Companies must read them one by one to find suitable employees, but this process can be time-consuming and error-prone, potentially limiting the chance of getting the best candidates.

This research aims to develop a curriculum vitae extraction system using Named Entity Recognition (NER). This research creates a system that automatically identifies and classifies the contents of curriculum vitae using Named Entity Recognition (NER) with a Rule Based approach, a core component of information extraction that labels input data. This research trained a model to detect entities in curriculum vitae text, with 100% precision, 87% recall and 93% f-score.

Keywords: *Curriculum Vitae, Information Extraction, Recruitment, Named Entity Recognition (NER), Rule Based*