

**PEMBANGUNAN APLIKASI IDENTIFIKASI
TYPOGRAPHICAL ERROR PADA DOKUMEN
BERBAHASA INDONESIA MENGGUNAKAN
ALGORITMA JARO-WINKLER *DISTANCE***

Tugas Akhir

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat
Sarjana Teknik Informatika**



Dibuat Oleh:

GRELLY LUCIA YOVELLIA LONDO
15 07 08471

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2019**

LEMBAR PENGESAHAN

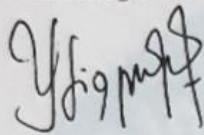
Pembangunan Aplikasi Identifikasi *Typographical Error* Pada Dokumen
Berbahasa Indonesia Menggunakan Algoritma Jaro-Winkler Distance
Yogyakarta, 26 Agustus 2019

Grelly Lucia Yovellia Londo

15 07 08471

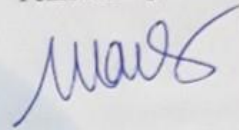
Menyetujui,

Pembimbing I



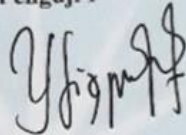
Y. Sigit Purnomo WP., S.T., M.Kom.
NPP: 04.01.710

Pembimbing II



Martinus Maslim, S.T., M.T.
NPP: 01.13.847

Penguji I



Y. Sigit Purnomo WP., S.T., M.Kom.
NPP: 04.01.710

Penguji II



Joseph Eric Samodra, S.Kom., MIT.
NPP: 04.15.896

Penguji III



Th. Adi Purnomo Sidhi, S.T., M.T.
NPP: 02.11.817

Mengetahui,

Dekan Fakultas Teknologi Industri


UNIVERSITAS SEBELA MARET
FAKULTAS TEKNOLOGI INDUSTRI

Dr. A. Teguh Siswanto, M.Sc.
NPP: 09.93.464

Pernyataan Orisinalitas & Publikasi Ilmiah

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Grelly Lucia Yovellia Londo
NPM : 15 07 08471
Program Studi : Teknik Informatika
Fakultas : Teknologi Industri
Judul Penelitian : Pembangunan Aplikasi Identifikasi Typographical Error
Pada Dokumen Berbahasa Indonesia Menggunakan
Algoritma Jaro Winkler *Distance*

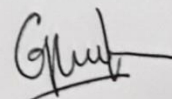
Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 13 Agustus 2019

Yang menyatakan,



Grelly Lucia Yovellia Londo

15 07 08471

HALAMAN PERSEMBAHAN

“Do your best and let God do the rest”.

– Ben Carson

Ia membuat segala sesuatu indah pada waktunya, bahkan Ia memberikan kekekalan dalam hati mereka. Tetapi manusia tidak dapat menyelami pekerjaan yang dilakukan Allah dari awal sampai akhir.

Pengkhotbah 3:11

Tugas akhir ini saya persembahkan untuk

Tuhan Yesus Kristus,

Papa, Mama, Oma, Opa

Keluarga

Sahabat-sahabat tercinta

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa karena atas tuntunan dan penyertaan-Nya penulis dapat menyelesaikan pembuatan tugas akhir dengan judul “Pembangunan Aplikasi Identifikasi Typographical Error Pada Dokumen Berbahasa Indonesia Menggunakan Algoritma Jaro-Winkler *Distance*” dengan baik.

Penulisan tugas akhir ini dilakukan untuk memenuhi salah satu syarat untuk mencapai derajat Sarjana Komputer dari Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam penulisan tugas akhir ini tidak terlepas dari bimbingan, dukungan serta bantuan banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

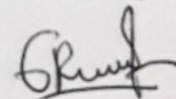
1. Tuhan Yang Maha Esa yang selalu membimbing, memberikan berkat-Nya, dan menyertai penulis.
2. Bapak Dr. A. Teguh Siswantoro, M.Sc., selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
3. Bapak Martinus Maslim, S.T., M.T. selaku Kepala Program Studi Teknik Informatika Universitas Atma Jaya Yogyakarta dan juga sebagai dosen pembimbing II yang meluangkan waktu dan pikiran untuk membimbing, memberikan ide dan masukan dalam menyelesaikan tugas akhir ini.
4. Bapak Y. Sigit Purnomo WP., S.T., M.Kom. selaku dosen pembimbing I yang telah meluangkan waktu dan pikiran untuk membimbing, memberikan ide, masukan serta bantuan dalam menyelesaikan tugas akhir ini.
5. Ibu Stephanie Pamela Adithama, S.T., M.T. selaku dosen pembimbing akademik yang telah membimbing penulis selama menempuh Pendidikan di Program Studi Teknik Informatika, Universitas Atma Jaya Yogyakarta.

6. Papa, Mama, Opa, Oma, Ka Sandra, Jovi, Gladys dan keluarga tercinta, yang memberikan doa dan dukungan baik moral maupun materi kepada penulis selama ini dari awal hingga akhir.

7. Egik yang selalu mendengar keluh kesah penulis, mendoakan, menghibur serta memberikan semangat.
8. Gege, Dea, Fara, Rani, Megan yang telah menjadi sahabat dan saudara selama masa perkuliahan yang selalu membantu dan menyemangati.
9. Melita, Anggy, Tiur, Rizka, Indah, Verry, Dege, Andreas, Dio, Roni, Hendy, Bonar, Tian, Ian, Alan, Wahyu, Eka, Sammy, Daniel, Kevin, Elen dan lainnya yang tidak dapat disebut satu persatu yang telah menjadi sahabat yang baik selama masa perkuliahan di UAJY yang selalu mendukung, membantu dan memberikan semangat kepada penulis.
10. Yayuk, Lidya, Nevel, Alisya, Mariana sebagai sahabat terbaik penulis yang selalu mendoakan dan memberikan semangat dari jauh.
11. Diovina, Lady, Lilian, Risty sebagai sahabat sekaligus saudara selama merantau di Jogja yang selalu memberikan dukungan dan semangat bagi penulis selama menyelesaikan tugas akhir ini.
12. Grup bimbingan Pak Sigit yang selalu membantu dan memberikan informasi selama proses pengerjaan tugas akhir ini.
13. Teman-teman HIMAFORKA dan KSP yang telah menemani dan berproses bersama selama masa perkuliahan.
14. Semua pihak yang tidak dapat disebutkan satu persatu yang secara langsung maupun tidak langsung telah membantu penulis selama pengerjaan tugas akhir ini serta memberikan dukungan dan semangat.

Demikian laporan tugas akhir ini dibuat, penulis menyadari bahwa tugas akhir ini jauh dari sempurna. Maka dari itu penulis memohon maaf sebesar-besarnya apabila terdapat sesuatu yang kurang berkenan dan tidak sesuai. Semoga tugas akhir ini dapat bermanfaat bagi pembaca.

Yogyakarta, 13 Agustus 2019



Grely Lucia Yovellia Londo

15 07 08471

INTISARI

PEMBUATAN APLIKASI IDENTIFIKASI *TYPOGRAPHICAL ERROR* PADA DOKUMEN BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA JARO-WINKLER *DISTANCE*

Grelly Lucia Yovellia Londo

15 07 08471

Teks merupakan salah satu media yang digunakan manusia untuk berkomunikasi dan berinteraksi setiap hari terutama dalam bidang pendidikan misalnya pada penulisan laporan tugas akhir. Hal yang paling sering terjadi dalam pembuatan teks adalah kesalahan penulisan atau typographical error. Berdasarkan masalah tersebut diperlukan suatu aplikasi yang dapat membantu penulis agar dapat mengidentifikasi typographical error pada dokumen yang ditulisnya.

Aplikasi dibuat menggunakan Laravel versi 5.8 untuk tampilan web dan python versi 3 untuk pemrosesan dataset serta pembuatan model dan web service nya. Pembuatan model menggunakan library nltk sedangkan penerapan algoritma Jaro-Winkler distance menggunakan library pylibjaro. Dataset yang digunakan adalah opensource dataset berupa daftar kata dari KBBI yang diambil dari github Jim Geovedi. Aplikasi ini hanya mendukung file pdf dan doc atau docx.

Hasil dari model yang sudah dibuat diterapkan pada webservice dengan keluaran berupa data JSON yang berisi daftar kata yang bernilai benar atau salah, jumlah kata dokumen, jumlah kata yang benar, jumlah kata yang salah dan waktu eksekusi program. Webservice ini akan digunakan pada aplikasi web yang dibangun.

Kata kunci: NLP, Bahasa Indonesia, typographical error, Jaro-Winkler distance, spelling checker

Dosen Pembimbing I : Y. Sigit Purnomo WP., S.T., M.Kom.

Dosen Pembimbing II : Martinus Maslim, S.T., M.T.

Jadwal Sidang Tugas Akhir : 22 Agustus 2019

DAFTAR ISI

Judul	i
LEMBAR PENGESAHAN	ii
Pernyataan Orisinalitas & Publikasi Ilmiah	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1. Penelitian Terdahulu.....	5
2.2. Landasan Teori	12
2.2.1. <i>Natural Language Processing (NLP)</i>	12
2.2.2. <i>Typographical Error</i>	12
2.2.3. <i>Jaro-Winkler Distance</i>	13
2.2.4. <i>Spelling Checker</i>	14
BAB III METODOLOGI PENELITIAN	15
3.1. Sumber Data	15
3.2. Alat Penelitian	15
3.2.1. Perangkat Keras	15
3.2.2. Perangkat Lunak.....	16
3.3. Langkah-Langkah Penelitian.....	17
3.3.1. Studi Pustaka.....	18
3.3.2. Pengumpulan Bahan.....	18

3.3.3.	Analisis Dataset.....	18
3.3.4.	Pembuatan Model.....	18
3.3.5.	<i>Preprocessing</i> Data	19
3.3.6.	Evaluasi Model.....	19
3.3.7.	Pembuatan <i>Web Service</i>	19
3.3.8.	Pembuatan Aplikasi <i>Web</i>	19
3.3.9.	Pengujian Aplikasi <i>Web</i>	19
BAB IV ANALISIS DAN PEMBAHASAN		20
4.1.	<i>Preprocessing</i> data	20
4.2.	Hitung Manual Jaro-Winkler <i>Distance</i>	21
4.3.	Implementasi Algoritma Jaro Winkler <i>Distance</i> pada Model.....	23
4.4.	Implementasi <i>Parallel Processing</i> pada Model	24
4.5.	Evaluasi Model.....	27
BAB V IMPLEMENTASI MODEL PADA <i>WEB SERVICE</i> DAN APLIKASI <i>WEB</i>		39
5.1.	<i>Web Service</i>	39
5.2.	Aplikasi <i>Web</i>	46
BAB VI PENUTUP		49
6.1.	Kesimpulan.....	49
6.2.	Saran	49
Daftar Pustaka		50
LAMPIRAN.....		52

DAFTAR GAMBAR

Gambar 3.1. Diagram langkah penelitian	17
Gambar 4.2. Simulasi algoritma Jaro-Winkler distance di Jupyter Notebook.....	22
Gambar 4.3. Kode fungsi implementasi library Jaro-Winkler distance.....	23
Gambar 4.4. Proses multiprocessing pada kode Python	24
Gambar 4.5. Implementasi multiprocessing pada model menggunakan Python	25
Gambar 4.6. Sebagian isi dataset	35
Gambar 5.1. Implementasi model pada web service.....	39
Gambar 5.2. Hasil <i>deployment</i> di <i>localhost</i>	40
Gambar 5.3. Alur proses aplikasi web	46
Gambar 5.4. Tampilan awal aplikasi <i>web</i> bagian unggah file	47
Gambar 5.5. Hasil pengecekan <i>typo</i> menggunakan aplikasi <i>Typocheck</i>	48

DAFTAR TABEL

Tabel 2.1. Tabel pembanding dengan penelitian sebelumnya	7
Tabel 4.1. Tabel spesifikasi komputer untuk uji coba parallel processing	26
Tabel 4.2. Perbandingan waktu eksekusi parallel processing pada tiga komputer	27
Tabel 4.3. Tabel hasil evaluasi model.....	27
Tabel 4.4. Tabel perbandingan eksekusi model antara proses serial dan paralel..	36

