

**IMPLEMENTASI METODE NAZIEF & ADRIANI
MENGUNAKAN METODE N-GRAM DAN SORENSEN DICE
COEFFICIENT DALAM PENGECEKAN TYPOGRAPHY
PADA KATA BAHASA INDONESIA**

SKRIPSI



Oleh

DWI HASTA ADHI BHIRAWA

NIM 14 622 030

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH GRESIK**

2019

LEMBAR PENGESAHAN

SKRIPSI

IMPLEMENTASI METODE NAZIEF & ADRIANI MENGUNAKAN METODE N-GRAM DAN SORENSEN DICE COEFFICIENT DALAM PENGECEKAN TYPOGRAPHY PADA KATA BAHASA INDONESIA

Oleh

DWI HASTA ADHI BHIRAWA

NIM 14 622 030

Telah dipertahankan di depan Tim Penguji

Pada tanggal *10 juli 2019*

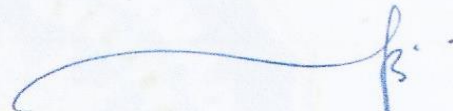
Susunan Tim Penguji

Penguji I (Ketua)

Penguji II (Sekretaris)

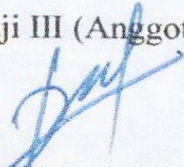


Deni Sutaji, S.Kom., M.Kom.
NIP: 06 211 211 153



Putri Aisyiyah Rakhma Devi, S.pd., M.Kom
NIP: 06 211 602 198

Penguji III (Anggota)



Henny Dwi Bhakti, S.Si., M.Si.
NIP: 06 211 709 200

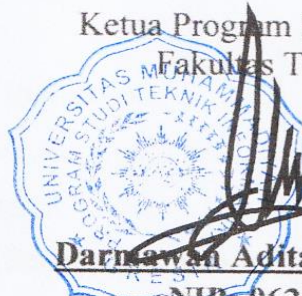
Diterima dan dinyatakan lulus sidang skripsi pada tanggal 16 juli 2019

Dekan Fakultas Teknik



H. H. Leksono, ST., MT., IPM
NIP: 197311122005011001

Ketua Program Studi Informatika
Fakultas Teknik UMG



Darnawan Aditama, S.Kom. M.T.
NIP: 0621 1602 197

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IMPLEMENTASI METODE NAZIEF & ADRIANI MENGUNAKAN METODE N-GRAM DAN SORENSEN DICE COEFFICIENT DALAM PENGECEKAN TYPOGRAPHY PADA KATA BAHASA INDONESIA

yang saya buat untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Gresik, sejauh yang saya ketahui bukan merupakan tiruan atau duplikasi dari skripsi yang sudah dipublikasikan dan atau pernah dipakai untuk mendapatkan gelar kesarjanaan di lingkungan Universitas Muhammadiyah Gresik maupun di Perguruan Tinggi atau Instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Pernyataan ini saya buat dengan sebenarnya dan apabila ternyata di kemudian hari saya terbukti melanggar pernyataan saya tersebut di atas, saya bersedia menerima sanksi sesuai aturan yang berlaku.

Gresik, 16 Juli 2019



DWI HASTA ADHI BHIRAWA

NIM 14 622 030

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGASAKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Muhammadiyah Gresik, saya yang bertandatangan di bawah ini :

Nama : Dwi Hasta Adhi Bhirawa
NIM : 14622030
Email : hastanobel44@gmail.com
Program Studi : Informatika
Fakultas : Teknik
Jenis karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Gresik Hak Bebas Royalti Noneksklusif (Non- exclusive RoyaltyFree Right) atas karya ilmiah saya yang berjudul :

IMPLEMENTASI METODE NAZIEF & ADRIANI MENGGUNAKAN METODE N-GRAM DAN SORENSEN DICE COEFFICIENT DALAM PENGECEKAN TYPOGRAPHY PADA KATA BAHASA INDONESIA beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas RoyaltiNoneksklusif ini Universitas Muhammadiyah Gresik berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir untuk kepentingan akademis tanpa perlu meminta-ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Muhammadiyah Gresik, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di :
Gresik
Pada tanggal :
26 Agustus 2024
Yang menyatakan



Dwi Hasta Adhi Bhirawa

Karya Ilmiah: Tugas Akhir

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan proposal skripsi dengan judul “Implementasi Metode Nazief & Adriani Menggunakan Metode N-Gram dan *Sorensen Dice Coefficient* Dalam Pengecekan *Typography* Pada Kata Bahasa Indonesia” Laporan proposal skripsi ini digunakan sebagai persyaratan menjadi Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Gresik, Untuk itu dalam penyusunan laporan skripsi ini penulis ingin mengucapkan terima kasih kepada:

1. Ibu dan Bapak yang ikhlas selalu mendoakan agar mudah dalam mengerjakan skripsi serta keluarga yang selalu mensupport penulis.
2. Bapak Indra Gita Anugrah, S.Kom., M.Kom selaku dosen pembimbing yang senantiasa meluangkan waktunya untuk memberikan bantuan, dukungan, arahan serta masukan bagi penulis dalam menyelesaikan skripsi.
3. Teman kelas A Sore, B Sore 2014 serta kakak-kakak Senior yang banyak membantu.
4. Teman-teman para pejuang skripsi Ima, Eka, Roni, Umam, Syukron, Habibi, Ucup, dll terutama Teknik Informatika angkatan 2014.
5. Semua pihak yang telah membantu penulis yang tidak dapat disebutkan namanya satu persatu oleh penulis

Penulis menyadari dalam penyusunan skripsi ini masih banyak kekurangan untuk itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca untuk perbaikan dimasa mendatang .

Gresik, 18 Juli 2019

Penulis,

IMPLEMENTASI METODE NAZIEF & ADRIANI MENGUNAKAN METODE N-GRAM DAN SORENSEN DICE COEFFICIENT DALAM PENGECEKAN TYPOGRAPHY PADA KATA BAHASA INDONESIA

Oleh

DWI HASTA ADHI BHIRAWA

14 622 030

Diajukan kepada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Gresik pada tanggal 15 Maret 2018 untuk memenuhi sebagian persyaratan memperoleh derajat sarjana S-1 Program Studi Teknik Informatika

INTISARI

Information Retrieval adalah studi tentang bagaimana data dapat ditemukan dari berbagai sumber atau kumpulan informasi relevan yang sedang dicari atau dibutuhkan. Ini melibatkan tindakan seperti pengindeksan, pencarian, dan pemanggilan kembali data, yang juga dikenal sebagai *information retrieval*. Data dapat mencakup teks, tabel, gambar, video, dan audio. Dalam *Information Retrieval*, *preprocessing* merupakan langkah penting, karena mengubah data input menjadi format yang sesuai dan siap untuk diproses. Metode stemming menggunakan algoritma Nazief & Adriani mematuhi kaidah morfologi bahasa Indonesia, mengklasifikasikan afiks ke dalam kategori seperti afiks yang diizinkan atau tidak diizinkan. Klasifikasi ini mencakup prefiks, sufiks, infiks, dan kombinasi afiks di awal dan akhir kata dasar (misalnya, dalam "membenarkan," yang mencakup afiks -me dan -kan yang diterapkan pada kata dasar "benar"). Penelitian ini akan difokuskan pada stemming kata-kata berafiks menggunakan metode Nazief & Adriani yang dikombinasikan dengan metode N-Gram dan *Sorensen Dice Coefficient*. Dalam pendekatan ini, untuk setiap kata yang kata dasarnya atau bentuk dasarnya tidak teridentifikasi, nilai N-Gram (Uni-Gram, Bi-Gram, Tri-Gram) akan dihitung untuk setiap suku kata dan dibandingkan dengan kamus Bahasa Indonesia. Setelah memperoleh nilai N-Gram, langkah selanjutnya adalah menghitung kedekatan atau jarak menggunakan Koefisien Sorensen-Dice (SDC). Langkah terakhir adalah mencari jarak terdekat ke suatu nilai konstan antara 0 dan 1, dengan kata yang memiliki nilai SDC tertinggi diasumsikan sebagai kata dasar, yang menunjukkan relevansi hasil.

Kata Kunci : *Information Retrieval*, *Stemming*, Nazief dan Adriani, *Sorensen Dice Coefficient*.

Pembimbing : Indra Gita Anugrah, S.Kom., M.Kom.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN SAMPUL LUAR	ii
HALAMAN SAMPUL DALAM	iii
LEMBAR PERNYATAAN.....	iv
LEMBAR PERSETUJUAN.....	v
LEMBAR PENGESAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
ABSTRAK.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah.....	3
1.5. Manfaat	3
1.6. Metodologi Penelitian	3
1.7. Sistematika Penulisan.....	4
BAB II LANDASAN TEORI	
2.1. Morfologi Bahasa Indonesia	6
2.2. Preprocessing.....	8
2.2.1. <i>Case Folding</i>	8
2.2.2. <i>Tokenizing</i> (Pemisahan rangkaian kata).....	8
2.2.3. <i>Penyaringan (Stopword removal)</i>	8
2.2.4. <i>Stemming</i>	9

2.2.5. <i>Stemming</i> Nazief & Adriani.....	9
2.3. N-Gram Metode.....	16
2.4. Sorensen Dice Coefficient.....	18
2.5. Penelitian Sebelumnya.....	19

BAB III ANALISIS DAN PERANCANGAN SISTEM

3.1. Analisis Sistem	20
3.2. Hasil Analisis.....	20
3.3. Perancangan Sistem.....	21
3.3.1. <i>Flowchart</i> Sistem.....	21
3.3.1.1. Kata Imbuhan.....	22
3.3.1.2. <i>Preprocessing</i>	22
3.3.1.2.1. <i>Case Folding</i>	23
3.3.1.2.2. <i>Tokenizing</i>	23
3.3.1.2.3. <i>Stopword</i>	24
3.3.1.2.1. <i>Stemming</i>	24
3.3.1.3. <i>Overstemming</i> atau <i>Understemming</i>	28
3.3.1.4. Hitung Jarak Kedekatan.....	29
3.3.1.5. <i>Similiarity</i> mendekati 1 atau <i>similiarity</i> = 1....	35
3.3.1.6. Kata Dasar.....	28
3.3.2. Diagram Konteks.....	37
3.3.3. Diagram Berjenjang.....	38
3.3.4. Data Flow Diagram (DFD).....	40
3.3.4.1. DFD level 0.....	40
3.3.4.2. DFD level 1.....	41
3.4. Perancangan Basis Data.....	42
3.4.1. Desain Tabel.....	42
3.4.2. Entity Relationship Diagram.....	45
3.5. Perancangan Antar Muka.....	46
3.5.1. Halaman Beranda.....	47
3.5.2. Halaman Kamus.....	47

3.5.3. Halaman Tambah Kamus.....	48
3.5.4. Halaman <i>Stemming</i>	49
3.5.5. Halaman N-Gram.....	50
3.6. Spesifikasi Pembuatan Sistem.....	50
3.7. Skenario Pengujian.....	51

BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Implementasi Sistem.....	53
4.1.1. Halaman Beranda	53
4.1.2. Halaman Kamus.....	54
4.1.3. Halaman <i>Stemming</i>	57
4.1.4. Halaman Hasil Proses SDC.....	62
4.2. Pengujian Sistem.....	64
4.2.1. Halaman Beranda	64
4.2.2. Halaman Kamus.....	64
4.2.3. Halaman <i>Stemming</i>	65
4.3. Analisis Hasil Pengujian Sistem.....	66
4.6. Pengujian Akurasi.....	67

BAB V PENUTUP

5.1. Kesimpulan	69
5.2. Saran	70

DAFTAR PUSTAKA

DAFTAR GAMBAR

Gambar 3.1. Gambaran Umum Sistem.....	21
Gambar 3.2. Flowchart Tahap <i>Preprocessing</i>	22
Gambar 3.3. Alur Proses <i>Stemming</i>	25
Gambar 3.4. Diagram Konteks Modifikasi Nazief & Adriani	37

Gambar 3.5.	Diagram Berjenjang Modifikasi Nazief & Adriani	38
Gambar 3.6.	Data Flow Diagram Level 0	40
Gambar 3.7.	Data Flow Diagram Level 1 Proses 1	41
Gambar 3.8.	Data Flow Diagram Level 1 Proses 2	42
Gambar 3.9.	RD Nazief & Adriani	46
Gambar 3.10.	Tampilan Rancangan Halaman Beranda	47
Gambar 3.11.	Tampilan Rancangan Halaman Kamus	48
Gambar 3.12.	Tampilan Rancangan Halaman Tambah Kamus.....	48
Gambar 3.13.	Tampilan Rancangan Halaman <i>Stemming</i>	49
Gambar 3.14.	Tampilan Rancangan Halaman N-Gram	50
Gambar 4.1.	Halaman Beranda.....	53
Gambar 4.2.	Halaman Kamus.....	54
Gambar 4.3.	Kolom Cari Kata.....	55
Gambar 4.4.	Fungsi Hapus Data Kamus.....	56
Gambar 4.5.	Halaman Tambah Data Kamus.....	56
Gambar 4.6.	Halaman <i>Stemming</i>	57
Gambar 4.7.	Halaman Hasil Proses SDC.....	62
Gambar 4.8.	Halaman Beranda.....	64
Gambar 4.9.	Halaman Kamus.....	64
Gambar 4.10.	Kolom Masukkan Kata.....	65
Gambar 4.11.	Kolom Hasil Proses SDC.....	65

DAFTAR TABEL

Tabel 2.1.	Kombinasi Awalan-Akhiran yang tidak diizinkan.....	10
Tabel 2.2.	Aturan Pemenggalan Awalan <i>Stemmer</i> Nazief & Adriani.....	11
Tabel 2.3.	Cara Menentukan Tipe Awalan Untuk awalan “te-“	13
Tabel 2.4.	Jenis Awalan Berdasarkan Tipe Awalannya.....	13

Tabel 3.1.	Contoh Proses <i>Case Folding</i>	23
Tabel 3.2.	Contoh Proses <i>Tokenization</i>	24
Tabel 3.3.	Contoh Proses <i>Stopword</i>	24
Tabel 3.4.	Kombinasi Awalan-Akhiran Yang Tidak diizinkan.....	26
Tabel 3.5.	Contoh <i>Stemming</i> kata “membeku”.....	27
Tabel 3.6.	Tabel Kamus Kata Dasar.....	30
Tabel 3.7.	Nilai Gram Pada <i>String</i> 1 dan 2.....	31
Tabel 3.8.	Jumlah Gram sama (\cap) irisan.....	32
Tabel 3.9.	Nilai <i>Similarity</i>	34
Tabel 3.10.	Pengurutan Kata Dari Nilai Terbesar.....	36
Tabel 3.11.	Tabel Kata Dasar.....	43
Tabel 3.12.	Tabel Kata.....	43
Tabel 3.13.	Tabel <i>Preprocessing</i>	44
Tabel 3.14.	Tabel <i>Stemming</i>	44
Tabel 3.15.	Tabel N-Gram.....	45
Tabel 3.16.	Parameter Menghitung Akurasi.....	51
Tabel 3.17.	Tabel Kata Berimbuhan.....	51
Tabel 4.1.	Tabel Hasil Uji.....	51
Tabel 4.2.	Tabel Hasil Pengujian Sistem.....	51