



## **LAPORAN SKRIPSI**

# **IMPLEMENTASI DAN ANALISIS ALGORITMA STEMMING NAZIEF & ADRIANI DAN PORTER PADA DOKUMEN BER-BAHASA INDONESIA**

Disusun oleh :

Nama : Dwi Wahyudi  
NIM : 08.5.10113  
Jurusan : Teknik Informatika  
Jenjang Pendidikan : Strata 1

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**SINAR NUSANTARA**

**SURAKARTA**

**2016**



**(LAPORAN SKRIPSI)**

Laporan ini disusun guna memenuhi salah satu syarat untuk menyelesaikan

program pendidikan Strata 1

Pada

STMIK Sinar Nusantara Surakarta

Disusun oleh :

Nama : Dwi Wahyudi  
NIM : 08.5.10113  
Jurusan : Teknik Informatika  
Jenjang Pendidikan : Strata 1

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

**SINAR NUSANTARA**

**SURAKARTA**

**2016**



**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER  
SINAR NUSANTARA**

**SURAT PERNYATAAN PENULIS**

JUDUL : Implementasi Dan Analisis Algoritma Stemming Nazief &  
Adriani dan Porter pada Dokumen Berbahasa Indonesia  
NAMA : Dwi Wahyudi  
NIM : 08.5.10113

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain mengklaim bahwa skripsi ini sebagai karyanya yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta hak dan kewajiban yang melekat pada gelar tersebut”.

Surakarta, 19 Mei 2016

Penulis

Dwi Wahyudi

## **PERSETUJUAN LAPORAN SKRIPSI**

Nama pelaksana skripsi : Dwi Wahyudi  
Nomor Induk Mahasiswa : 08.5.10113  
Program Studi : Teknik Informatika  
Jenjang Pendidikan : Strata 1  
Judul Laporan Skripsi : Implementasi dan Analisis Algoritma  
Stemming Nazief & Adriani dan Porter  
Pada Dokumen ber-Bahasa Indonesia  
Dosen Pembimbing 1 : Teguh Susyanto, S.Kom, M.Cs  
Dosen Pembimbing 2 : Didik Nugroho, M.Kom

Surakarta, .....2016

Menyetujui

Dosen Pembimbing 1

Dosen Pembimbing 2

(Teguh Susyanto, S.Kom, M.Cs)

(Didik Nugroho, M.Kom)

Mengetahui

Ketua STMIK Sinar Nusantara Surakarta

(Kumaratih Sandradewi, S.P.,M.kom)

## MOTTO

- *“Barang siapa yang ingin meraih kesuksesan di dunia maka raihlah dengan ilmu, barang siapa yang ingin meraih kesuksesan di akhirat maka raihlah dengan ilmu, dan barang siapa yang ingin meraih kesuksesan keduanya maka raihlah dengan ilmu”.* (HR. Muslim)
- *“Jika seorang manusia mati maka terputuslah darinya amalnya kecuali dari tiga hal; sedekah jariyah, ilmu yang bermanfaat, anak shalih yang mendoakannya.”* (HR. Muslim no. 1631)
- *“Ikatlah ilmu dengan menuliskannya”* (Sayyidina Ali bin Abi Thalib)
- Manusia tidak akan menjadi bodoh dengan membagi ilmu, dan manusia tidak akan menjadi pandai dengan menyimpan ilmu
- Bahagia itu menerima dan bersyukur atas apa yang telah Allah SWT berikan
- Senyum itu indah dan mengindahakan

## **PERSEMBAHAN**

*Untuk Rabbku, Rasulku dan Agamaku*

Untuk keluargaku Babe, Mami, Mbak, *my real heroes*

Untuk *my little family*, istriku Yusi Firdausi, kesayangan ayah Alifah Hasna

Rosyidah dan Muhammad Al Fatih

Untuk para pejuang ilmu di bumi Allah azza wajalla

## SUMMERY

The thesis report with title “Implementasi dan Analisis Algoritma Stemming Nazief & Adriani dan Porter pada Dokumen ber-Bahasa Indonesia” was arranged on period September 2015 until March 2016.

The purpose of this thesis is to implemented and analyze of stemming algorithms of Nazief & Adriani and Porter on Indonesian documents to support performance of information retrieval system. The better one will be known by comparing the results of both of algorithms.

The method of collecting data are by download the test documents from web and study of literatures. The test documents are downloaded from online news website (detik.com) using crawler. The study of literatures is by search and collect all literatures all about Information Retrieval, stemming, Nazief & Adriani’s algorithm and Porter’s algorithm. For this thesis is refer to Fadillah Z Tala’s thesis and Jelita Asian’s thesis.

The both of stemmers are implemented on a web aplication that developed using PHP and MySQL. The result for each stemmer will be checked manually to determine the correct one. Final result shown that Nazief & Adriani’s algorithm is better than Porter’s algorithm for support information retrieval performance. Other side, Porter’s algorithm is better for procesing time.

## RINGKASAN

Laporan skripsi dengan judul Implementasi dan Analisis Algoritma Stemming Nazief & Adriani dan Porter pada Dokumen ber-Bahasa Indonesia telah dilaksanakan pada bulan September 2015 sampai dengan Maret 2016.

Tujuan dari penelitian ini adalah untuk mengimplementasikan dan menganalisa algoritma *stemming* Nazief & Adriani dan Porter pada dokumen atau teks ber Bahasa Indonesia untuk mendukung performa sistem temu kembali informasi (*Information Retrieval*). Dari hasil yang didapat, kemudian dibandingkan untuk mengetahui tingkat kesalahan dari kedua *stemmer* tersebut.

Metode pengumpulan data dilakukan dengan mengunduh dokumen uji dan studi kepustakaan. Dokumen uji diunduh dari situs berita *online* (detik.com) menggunakan *crawler*. Studi kepustakaan dilakukan dengan penelitian kepustakaan yang relevan dengan *Information Retrieval*, *stemming*, algoritma Nazief & Adriani dan algoritma Porter untuk stemming Bahasa Indonesia. Penelitian ini lebih mengacu pada penelitian yang dilakukan oleh Fadillah Z Tala dan Jelita Asian.

Kedua *stemmer* tersebut diimplementasikan pada sebuah aplikasi berbasis web yang dibangun menggunakan bahasa pemrograman PHP dan database MySQL. Hasil *stemming* dari kedua *stemmer* tersebut dianalisis secara manual untuk menentukan benar dan salahnya kata yang didapat. Hasil dari analisis menunjukkan bahwa algoritma Nazief & Adriani memberikan hasil yang lebih baik dibanding algoritma Porter untuk mendukung sistem temu balik informasi (*information retrieval*). Tetapi algoritma Porter memiliki waktu proses yang lebih cepat.



## KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Tuhan Yang Maha Esa yang telah memberikan segala nikmat dan karunia-Nya, sehingga tersusunlah Laporan Skripsi ini dengan judul “Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani dan Porter Pada Dokumen ber-Bahasa Indonesia”.

Penyusunan skripsi ini merupakan salah satu kewajiban yang dimaksud untuk menyelesaikan program pendidikan Strata 1 pada STMIK Sinar Nusantara, Surakarta.

Atas tersusunnya laporan skripsi ini, tidak lupa Penulis mengucapkan banyak terima kasih yang sebesar-besarnya kepada:

1. Ibu Kumaratih Sandra Dewi, selaku Ketua STMIK Sinar Nusantara Surakarta.
2. Bapak Teguh Susyanto, S.Kom, M.Cs selaku Pembimbing Akademik Penulis sekaligus Pembimbing I.
3. Bapak Didik Nugroho, M.Kom, selaku Pembimbing II.
4. Bapak/Ibu dosen dan karyawan STMIK Sinar Nusantara Surakarta.
5. Keluarga, sahabat, teman-teman dan semua pihak yang telah membantu tersusunnya laporan skripsi ini.

Penulis sangat berharap, semoga skripsi ini bisa bermanfaat.

Surakarta, April 2016

Penulis

## DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN .....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN MOTTO .....	v
HALAMAN PERSEMBAHAN .....	vi
RINGKASAN .....	viii
KATA PENGANTAR .....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xv
DAFTAR LAMPIRAN .....	xvi
BAB I : PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Kerangka Pikiran.....	4
1.7 Sistematika Penulisan.....	6
BAB II : LANDASAN TEORI.....	7
2.1 Sistem Temu Kembali (Information Retrieval) .....	7

2.2 Stemming .....	9
2.3 Algoritma Nazief & Adriani .....	12
2.4 Algoritma Porter.....	21
2.5 Algoritma Porter untuk Bahasa Indonesia .....	22
<b>BAB III : METODE PENELITIAN.....</b>	<b>28</b>
3.1 Pengumpulan Data .....	28
3.2 Analisis dan Perancangan Sistem.....	32
3.2.1 Kebutuhan Perangkat Lunak.....	33
3.2.2 Flowchart .....	33
3.2.3 Rancangan aplikasi .....	35
3.3 Analisis Hasil .....	36
<b>BAB VI : GAMBARAN UMUM.....</b>	<b>38</b>
4.1 Gambaran Umum Program .....	38
4.2 Analisis Algoritma Nazief & Adriani .....	38
4.3 Analisis Algoritma Porter .....	45
<b>BAB V : PEMBAHASAN .....</b>	<b>48</b>
5.1 Diagram Alir ( <i>Flowchart</i> ).....	48
5.2 Implementasi .....	51
5.2.1 Persiapan Data .....	51
5.2.2 Dokumen Uji .....	54
5.2.3 Algoritma Porter .....	56
5.2.4 Algoritma Nazief & Adriani.....	61
5.3 Analisis Hasil .....	66

BAB VI : PENUTUP .....	69
6.1 Kesimpulan .....	69
6.2 Saran.....	70
DAFTAR PUSTAKA .....	72
LAMPIRAN.....	74

## DAFTAR TABEL

Tabel 2.1 Kombinasi awalan dan akhiran yang tidak diijinkan.....	16
Tabel 2.2 Aturan pemenggalan awalan.....	17
Tabel 2.3 Penambahan aturan pemenggalan awalan.....	20
Tabel 2.4 Perubahan aturan pemenggalan awalan .....	20
Tabel 2.5 Kelompok <i>rule</i> pertama. <i>Inflection Particle</i> .....	24
Tabel 2.6 Kelompok <i>rule</i> kedua. <i>Inflectional Possesive Pronouns</i> .....	24
Tabel 2.7 Kelompok <i>rule</i> ketiga. <i>First order of derivational prefixes</i> .....	25
Tabel 2.8 Kelompok <i>rule</i> keempat. <i>Second order of derivaritional prefixes</i> ..	25
Tabel 2.9 Kelompok <i>rule</i> kelima. <i>Derivarional suffixes</i> .....	26
Tabel 2.10 <i>Illegal confix pairs</i> algoritma Porter .....	26
Tabel 2.11 Urutan awalan ganda yang diperbolehkan.....	27
Tabel 3.1 Daftar Tanda Baca .....	29
Tabel 5.1 Daftar indeks dokumen uji.....	51
Tabel 5.2 Jumlah kata, <i>doublewords</i> , <i>stopwords</i> dan <i>stemming words</i> .....	55
Tabel 5.3 Hasil Pengujian <i>stemming</i> algoritma Porter .....	56
Tabel 5.4 Contoh kata <i>overstemming</i> pada algoritma Porter .....	58
Tabel 5.5 Contoh kata <i>understemming</i> pada algoritma Porter.....	59
Tabel 5.6 Contoh kata <i>unchange</i> pada algoritma Porter.....	60
Tabel 5.7 Hasil pengujian <i>stemming</i> pada algoritma Nazief & Adriani.....	61
Tabel 5.8 Contoh kata <i>unchange</i> pada algoritma NA.....	63
Tabel 5.9 Contoh kata <i>overstemming</i> pada algoritma NA.....	64
Tabel 5.10 Contoh kata <i>understemming</i> pada algoritma NA.....	65

Tabel 5.11 Contoh perbandingan hasil algoritma NA dan Porter .....	66
Tabel 5.12 Contoh kesalahan <i>stemming</i> Nazief & Adriani dan Porter .....	67
Tabel 5.13 Perbandingan hasil <i>stemming</i> Porter dan Nazief & Adriani .....	68

## DAFTAR GAMBAR

Gambar 1.1 Skema pikiran implementasi dan analisis algoritma <i>stemming</i> ....	5
Gambar 2.1 Algoritma Porter untuk Bahasa Indonesia .....	23
Gambar 3.1 <i>Flowchart</i> implementasi algoritma <i>stemming</i> .....	34
Gambar 5.1 <i>Flowchart</i> unduh dokumen uji .....	48
Gambar 5.2 <i>Flowchart</i> proses <i>stemming</i> .....	50

## DAFTAR LAMPIRAN

Lampiran	Halaman
1. Daftar Stopword Tala.....	74