



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

**PERBANDINGAN PERFORMA ALGORITMA NAZIEF-
ADRIANI, SUCCESSOR VARIETY METODE CUTOFF, DAN
N-GRAM TERMODIFIKASI DALAM STEMMING DOKUMEN**

SKRIPSI

diajukan guna memenuhi sebagian persyaratan memperoleh gelar
Sarjana Komputer (S. Kom.)



Hadi Hartono
10110110069

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN KOMUNIKASI
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2015**

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN PERFORMA ALGORITMA NAZIEF- ADRIANI, SUCCESSOR VARIETY METODE CUTOFF, DAN N-GRAM TERMODIFIKASI DALAM STEMMING DOKUMEN

oleh

Nama : Hadi Hartono

NIM : 10110110069

Fakultas : ICT

Program Studi : Teknik Informatika

Tangerang, 18 Februari 2015

Ketua Sidang

Dosen Penguji

Maria Irmina P., S.Kom., M.T.

Nanang Krisdianto, S.T., M.Kom.

Dosen Pembimbing I

Dodick Z. Sudirman S.Kom., M.T.I.

Mengetahui,

Ketua Program Studi
Teknik Informatika

Maria Irmina P., S.Kom., M.T.

PERNYATAAN TIDAK MELAKUKAN PLAGIAT

Dengan ini saya :

Nama : Hadi Hartono
NIM : 10110110069
Fakultas : Teknologi Informasi dan Komunikasi
Program Studi : Teknik Informatika

menyatakan bahwa skripsi “Perbandingan Performa Algoritma *Stemming* Nazief-Adriani, *Successor Variety* Metode *Cutoff*, dan N-gram Termodifikasi dalam *Stemming* Dokumen” ini adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah yang ditulis oleh orang lain atau lembaga lain, dan semua karya ilmiah orang lain atau lembaga lain yang dirujuk dalam skripsi ini telah disebutkan sumber kutipannya serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah Skripsi yang telah saya tempuh.

Tangerang, 26 Januari 2015

Hadi Hartono

**PERBANDINGAN PERFORMA ALGORITMA NAZIEF-ADRIANI,
SUCCESSOR VARIETY METODE CUTOFF, DAN N-GRAM
TERMODIFIKASI DALAM STEMMING DOKUMEN**

ABSTRAKSI

Skripsi ini memaparkan perbandingan antara tiga jenis algoritma *stemming*, yaitu algoritma Nazief-Adriani, *Successor Variety* metode *Cutoff*, dan N-gram dalam *stemming* dokumen, khususnya dokumen berbahasa Indonesia. Algoritma-algoritma tersebut digunakan untuk mendapatkan *stem* (kata dasar) dari kata-kata yang diproses dimana *stem* tersebut dapat digunakan untuk meningkatkan performa dalam proses temu kembali informasi. Parameter pengujian menggunakan dua parameter dari metode evaluasi Paice, yaitu *Understemming Index* (UI) dan *Overstemming Index* (OI), dimana makin mendekati nilai nol, makin tinggi pula tingkat akurasi *stem* yang dihasilkan. Pengujian terhadap lima dokumen berbeda menyimpulkan algoritma Nazief-Adriani adalah algoritma dengan tingkat akurasi paling tinggi di antara ketiga algoritma yang diujikan dalam *stemming* dokumen berbahasa Indonesia.

Kata kunci : *Cutoff*, Nazief-Adriani, N-gram, Paice, *stemming*, *Successor Variety*.

**DOCUMENT-STEMMING PERFORMANCE COMPARISON BETWEEN
NAZIEF-ADRIANI, SUCCESSOR VARIETY WITH CUTOFF METHOD,
AND MODIFIED N-GRAM ALGORITHMS**

ABSTRACT

This essay explains a comparison between three stemming algorithm, namely Nazief-Adriani algorithm, Successor Variety with Cutoff method, and N-gram to stem Indonesian documents. Stems derived from these algorithms can be used to improve information retrieval performance. The criterion used to compare these algorithms is taken from Paice evaluation method, namely understemming index and overstemming index. The closer those indexes to zero means the higher accuracy of stems derived from these algorithms. Testing made to five different documents concluded that Nazief-Adriani algorithm is the most accurate stemmer among compared algorithms to stem Indonesian documents.

Keywords : Cutoff, Nazief-Adriani, N-gram, Paice, stemming, Successor Variety.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya skripsi berjudul “Perbandingan Performa Algoritma Nazief-Adriani, *Successor Variety* metode *Cutoff* dan N-gram Termodifikasi dalam *Stemming* Dokumen” ini dapat diselesaikan.

Terselesaikannya skripsi ini tidak lepas dari kerja sama banyak pihak. Ijinkan penulis mengucapkan terima kasih kepada:

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara;
2. Maria Irmina P., S.Kom., M.T., selaku Kaprodi Teknik Informatika;
3. Dodick Sudirman S.Kom., M.T.I., selaku dosen pembimbing, yang telah meluangkan waktunya dan memberikan banyak masukan mengenai penyusunan laporan ini;
4. Orang tua dan keluarga tersayang yang selalu memberikan dukungan dan semangat dalam segala hal;
5. Paulus M. Bodhi dan Destia Wijaya, teman seangkatan yang telah banyak memberikan dukungan dan dorongan selama masa perkuliahan;
6. Dosen-dosen yang telah berbagi ilmu dan pengetahuan selama masa perkuliahan yang telah dilalui, dan
7. Semua pihak yang lupa disebutkan atau tak bisa disebutkan satu persatu yang secara tidak langsung telah membantu penulis menyelesaikan skripsi ini.

Dalam penyusunan skripsi ini pasti banyak kekurangan. Untuk itu saran dan kritik sangat diharapkan, dengan harapan skripsi ini dapat memberikan manfaat bagi kita semua.

Tangerang, November 2014

Penulis

DAFTAR ISI

PERNYATAAN TIDAK MELAKUKAN PLAGIAT	i
ABSTRAKSI	ii
ABSTRACT	iii
KATA PENGANTAR	iv
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR RUMUS	ix
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 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 Information Retrieval	6
2.2 Stemming	7
2.3 Algoritma Nazief-Adriani	7
2.4 Algoritma Successor Variety	13
2.5 Algoritma N-gram	17
2.6 Algoritma N-gram Termodifikasi	17
2.7 Metode Evaluasi Paice	20
BAB III METODOLOGI DAN PERANCANGAN SISTEM	25
3.1 Metode Penelitian	25
3.2 Perancangan Sistem	26
3.2.1 Flow Chart	26
3.2.2 Data Flow Diagram	32
3.2.3 Struktur Tabel Basis Data	35
3.2.4 Desain Antarmuka Aplikasi	38
3.3 Pengembangan Sistem	40
3.3.1 Alat dan Bahan	41
3.3.2 Implementasi	42
BAB IV PENGUJIAN DAN PEMBAHASAN	49
4.1 Pengujian	49
4.1.1 Data	49

4.1.2	Kamus.....	50
4.1.3	Variabel Pengujian	50
4.1.4	Pengujian Fungsionalitas.....	50
4.1.5	Hasil Pengujian.....	56
BAB V KESIMPULAN DAN SARAN.....		62
5.1	Kesimpulan.....	62
5.2	Saran.....	62
DAFTAR PUSTAKA		63
DAFTAR LAMPIRAN.....		65

UMN

DAFTAR GAMBAR

Gambar 2.1 <i>Infomation Retrieval Process</i>	7
Gambar 3.1 <i>Flow chart</i> aplikasi.....	27
Gambar 3.2 <i>Flow chart</i> algoritma Nazief-Adriani	28
Gambar 3.3 <i>Flow chart</i> subproses cek kata ulang algoritma Nazief-Adriani.....	29
Gambar 3.4 <i>Flow chart</i> algoritma N-gram termodifikasi.....	30
Gambar 3.5 <i>Flow chart</i> algoritma <i>Successor Variety</i> metode <i>Cutoff</i>	31
Gambar 3.6 DFD Level 0 atau <i>Context Diagram</i> Aplikasi Perbandingan Performa Algoritma <i>Stemming</i>	32
Gambar 3.7 DFD Level 1 Aplikasi Perbandingan Performa Algoritma <i>Stemming</i>	33
Gambar 3.8 DFD Level 2 Proses <i>Stemming</i>	34
Gambar 3.9 Desain antarmuka <i>form</i> pemilihan dokumen.....	38
Gambar 3.10 Desain antarmuka <i>form</i> penentuan kelompok kata.....	39
Gambar 3.11 Desain antarmuka <i>form</i> penentuan benar-salahnya <i>stem</i>	39
Gambar 3.12 Desain antarmuka <i>form</i> kelompok kata berdasarkan hasil <i>stemming</i>	40
Gambar 3.13 Desain antarmuka <i>form</i> nilai UI dan OI.....	40
Gambar 3.14 Tampilan awal <i>form</i> pemilihan dokumen	42
Gambar 3.15 Tampilan ketika proses <i>stemming</i> sedang berlangsung	43
Gambar 3.16 Tampilan ketika proses <i>stemming</i> selesai	43
Gambar 3.17 Tampilan awal <i>form</i> penentuan kelompok kata	44
Gambar 3.18 Tampilan ketika pengguna selesai mengelompokkan kata.....	44
Gambar 3.19 Tampilan awal <i>form</i> penentuan benar-salahnya <i>stem</i>	45
Gambar 3.20 Tampilan ketika selesai menentukan benar-salahnya <i>stem</i>	46
Gambar 3.21 Dialog pemastian data telah diisi dengan benar	46
Gambar 3.22 Tampilan setelah aplikasi selesai menyimpan data	47
Gambar 3.23 Tampilan <i>form</i> kelompok kata berdasarkan hasil <i>stemming</i>	48
Gambar 3.24 Tampilan <i>form</i> hasil evaluasi.....	48
Gambar 4.1 Nilai UI dan OI Stefani.txt.....	60

UMN

DAFTAR TABEL

Tabel 2.1 Imbuhan pada bahasa Indonesia.....	12
Tabel 2.2 Contoh <i>Successor Variety</i>	14
Tabel 3.1 Entitas Aplikasi Perbandingan Performa Algoritma <i>Stemming</i>	33
Tabel 3.2 Aliran Data DFD Level 1 Aplikasi Perbandingan Performa Algoritma <i>Stemming</i>	33
Tabel 3.3 Aliran Data DFD Level 2 Proses <i>Stemming</i>	35
Tabel 3.4 Struktur tabel tb_katadasar awal	35
Tabel 3.5 Struktur tabel tb_katadasar akhir.....	36
Tabel 3.6 Struktur tabel tb_dokumen.....	36
Tabel 3.7 Struktur tabel tb_hasil.....	37
Tabel 3.8 Struktur tabel tb_kelompok.....	37
Tabel 3.9 Struktur tabel tb_isi_kelompok	37
Tabel 3.10 Struktur tabel tb_nilai	38
Tabel 4.1 Daftar dokumen yang diujikan	49
Tabel 4.2 <i>Successor value</i> dan <i>Predecessor value</i> kata “kerja”	51
Tabel 4.3 <i>Successor value</i> dan <i>Predecessor value</i> kata “bersekolah”	54
Tabel 4.4 <i>Semantic group</i> dan hasil <i>stemming</i>	57
Tabel 4.5 Nilai DMT, UMT, dan DNT <i>semantic group</i>	60
Tabel 4.6 Nilai UI dan OI Hasil Pengujian	61

DAFTAR RUMUS

Rumus 2.1.....	8
Rumus 2.2.....	15
Rumus 2.3.....	15
Rumus 2.4.....	18
Rumus 2.5.....	18
Rumus 2.6.....	18
Rumus 2.7.....	19
Rumus 2.8.....	19
Rumus 2.9.....	21
Rumus 2.10.....	22
Rumus 2.11.....	22
Rumus 2.12.....	22
Rumus 2.13.....	22
Rumus 2.14.....	23
Rumus 2.15.....	23
Rumus 2.16.....	24
Rumus 2.17.....	24
Rumus 2.18.....	24

UMN