

**ANALISIS PERBANDINGAN ALGORITMA STEMMING NAZIEF &
ADRIANI DENGAN ALGORITMA STEMMING PORTER UNTUK
STEMMING BAHASA INDONESIA**

SKRIPSI



disusun oleh

Ananda Febry Hartomi

16.11.0562

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**ANALISIS PERBANDINGAN ALGORITMA STEMMING NAZIEF &
ADRIANI DENGAN ALGORITMA STEMMING PORTER UNTUK
STEMMING BAHASA INDONESIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Ananda Febry Hartomi

16.11.0562

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA STEMMING NAZIEF &
ADRIANI DENGAN ALGORITMA STEMMING PORTER UNTUK
STEMMING BAHASA INDONESIA**

yang dipersiapkan dan disusun oleh

Ananda Febry Hartomi

16.11.0562

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal

Dosen Pembimbing,

Krisnawati, S.Si., M.T
NIK. 190302038

PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA STEMMING NAZIEF &
ADRIANI DENGAN ALGORITMA STEMMING PORTER UNTUK
STEMMING BAHASA INDONESIA**

yang dipersiapkan dan disusun oleh

Ananda Febry Hartomi

16.11.0562

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 November 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Erni Seniwati, S.Kom, M.Cs
NIK. 190302231

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Krisnawati, S.Si., M.T.
NIK. 190302038

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si., M.T
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 18 Februari 2021



Ananda Febry Hartomi

NIM. 16.11.0562

MOTTO

”Setiap hari kamu punya dua pilihan, melanjutkan tidur dan bermimpi atau bangun untuk mengejarnya”

(IngetUmur)

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan, Sesungguhnya sesudah kesulitan itu ada kemudahan.”

(QS. AL-Insyirah: 5-6)

”Berbahagialah wahai para tersepelekan, karena dengan begitu kalian punya kesempatan besar untuk mengejutkan”

(Faridstevy)

”Don't look back in anger”

(OASIS)

PERSEMBAHAN

Segala puji dan syukur kupersembahkan kepada Allah SWT yang Maha Kuasa yang tidak pernah meninggalkan dan mengabulkan doa yang selalu kupanjatkan. Terimakasih atas rasa syukur, nikmat, dan karunia yang telah Engkau berikan. Terimakasih Engkau telah memberiku pertolongan, kekuatan, kesabaran, ilmu, serta memberiku orang-orang di sekelilingku yang menyayangiku, selalu memberiku semangat dan doa sehingga skripsi ini dapat terselesaikan. Untuk itu kuucapkan rasa terimakasihku juga kepada:

1. Bapakku Hari Soeprpto dan Ibuku Sukapti, yang telah membesarkan aku, memberi kasih sayang tulus, mendidikku, memberi nasehat, motivasi, dukungan, doa, dan berjuang segalanya demi hidupku. Juga adikku Aditya Dwi Hardianti yang juga mendoakan aku dan membuat aku termotivasi untuk cepat lulus.
2. Dosen Pembimbing saya, Ibu Krisnawati, S.Si, MT yang telah sabar membimbing dan membantu dalam pengerjaan skripsi ini.
3. Sahabat-sahabat sekelas saya IF-09 berbagi suka dan duka bersama saya, serta sahabat main/nogkrong Yang telah menemani saya dari awal sampai ahir perkuliahan, khususnya “grub gendadapan”
4. Sahabat-sahabat Kos SEMUD, yang telah mau berbagi suka dan duka bersama saya dari awal sampai akhir masa perkuliahan saya, banyak sekali proses pendewasaan yang saya dapatkan disana, semoga kita bisa saling bertegur sapa Kembali
5. Terima kasih banyak untuk semua sistem pendukung yang tidak bisa saya sebutkan satu persatu, semoga sehat terus serta bahagia selalu lancar rezeki dimanapun kalian berada, terimakasih untuk empat tahunnya.

KATA PENGANTAR

Assalamualaikum Wr.Wb.

Alhamdulillah, segala puji syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan ridho-NYA, sehingga skripsi dengan judul “Analisis Perbandingan Algoritma Stemming Nazief & Adriani Dengan Algoritma Stemming Porter Untuk Stemming Bahasa Indonesia” ini dapat diselesaikan. Skripsi ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana pada program studi Informatika Universitas Amikom Yogyakarta.

Dengan selesainya skripsi ini, maka penulis tidak lupa mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku dekan Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
3. Bapak Sudarmawan, S.T., M.T. selaku Ketua Progam Studi Informatika.
4. Ibu Krisnawati, S.Si, M.T. selaku dosen pembimbing yang telah banyak memberikan arahan dalam pembuatan skripsi.
5. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah banyak memberikan ilmunya selama masa kuliah.

Peneliti tentunya menyadari bahwa pembuatan skripsi ini masih banyak kekurangan dan kelemahannya. Oleh karena itu peneliti berharap kepada semua pihak agar dapat menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan skripsi ini. Namun peneliti tetap berharap skripsi ini akan bermanfaat bagi semua pihak yang membacanya.

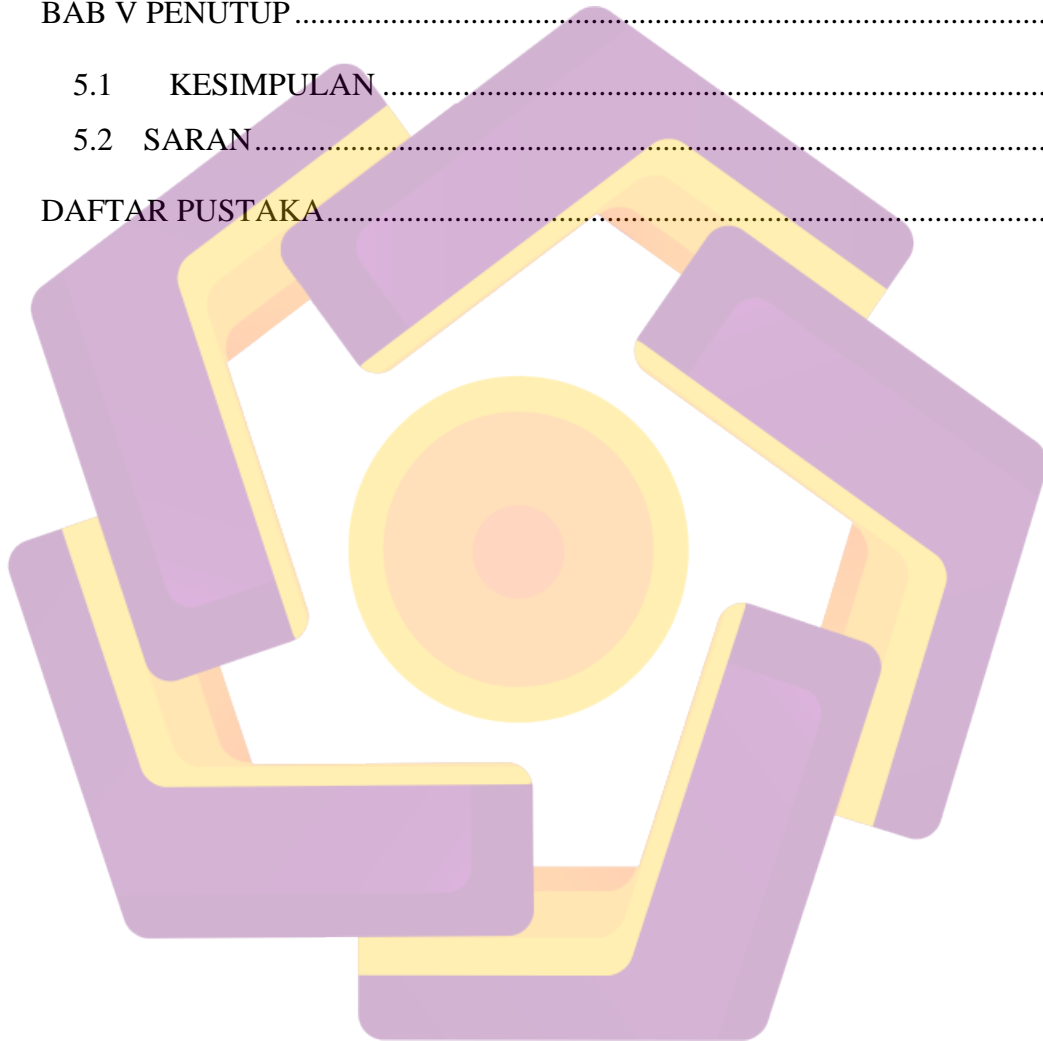
Wassalamualaikum Wr.Wb.

DAFTAR ISI

JUDUL	i
PERSETUJUAN SKRIPSI.....	II
PENGESAHAN	III
PERNYATAAN	iv
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR.....	VII
DAFTAR ISI	VIII
DAFTAR TABEL	XI
DAFTAR GAMBAR.....	XII
INTISARI	XIII
ABSTRACT	XIV
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	4
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN.....	5
1.7 SISTEMATIKA PENULISAN.....	7
1.7.1 BAB I : Pendahuluan.....	7
1.7.2 BAB II : Landasan Teori	7
1.7.3 BAB III : Analisis dan Perancangan.....	7
1.7.4 BAB VI : Implementasi dan Pembahasan	7
1.7.5 BAB V : Penutup.....	7
BAB II LANDASAN TEORI.....	8

2.1	TINJAUAN PUSTAKA	8
2.2	DASAR TEORI	11
2.2.1	Text Mining	11
2.2.2	Text Preprocessing	11
2.2.3	Case Folding	11
2.2.4	Tokenizing	12
2.2.5	Stop Word Filtering	12
2.2.6	Stemming	12
2.2.7	Parameter Analisis Hasil	22
2.2.8	Aplikasi Berbasis Web	23
2.2.9	TEORI TEKNIK ANALISA YANG DIGUNAKAN	26
BAB III METODE PENELITIAN		32
3.1	ANALISIS MASALAH	32
3.2	HASIL ANALISIS	32
3.3	SOLUSI YANG DAPAT DI TERAPKAN	33
3.4	ANALISA KEBUTUHAN	33
3.4.1	Kebutuhan Perangkat Keras	33
3.4.2	Kebutuhan Perangkat Lunak	34
3.4.3	Kebutuhan Fungsional	34
3.4.4	Kebutuhan Non Fungsional	35
3.5	ALUR PENELITIAN	35
3.5.1	Persiapan Data	36
3.5.2	Preprocessing Text	36
3.5.3	Penerapan Algoritma Stemming	38
3.5.4	Analisis Hasil	41
3.6	PERANCANGAN SYSTEM	42
3.6.1	Perancangan Proses	42
BAB IV ANALISIS DAN PEMBAHASAN		52
4.1	PERSIAPAN DATA	52
4.2	PREPROCESSING TEXT	53
4.3	PENERAPAN ALGORITMA	57

4.3.1	Algoritma Nazief&Adriani.....	57
4.3.2	Algoritma Porter.....	61
4.4	PARAMETER ANALISIS HASIL.....	65
4.4.1	Akurasi Word Conflation Class.....	65
4.4.2	Index Compression Factor.....	66
4.4.3	Akurasi waktu proses.....	67
BAB V PENUTUP.....		75
5.1	KESIMPULAN.....	75
5.2	SARAN.....	76
DAFTAR PUSTAKA.....		77



DAFTAR TABEL

Table 2.1	Table Perbandingan Penelitian	9
Table 2.2	Imbuhan Yang Tidak Diizinkan	15
Table 2.3	Aturan Penghapusan Awalan Me-	16
Table 2.4	Aturan Penghapusan Awalan Pe-	17
Table 2.5	Aturan Penghapusan Awalan Be-	17
Table 2.6	Aturan Penghapusan Awalan Te-	18
Table 2.7	Aturan Untuk Inflectional Particle	20
Table 2.8	Aturan Untuk Inflectional Possesive Pronoun	20
Table 2.9	Aturan Untuk First Order Derivational Prefix	21
Table 2.10	Aturan Untuk Second Order Derivational Prefix	21
Table 2.11	Aturan Untuk Derivational Suffix	22
Table 2.12	Simbol Flowchart	27
Table 2.13	Simbol DFD	29
Table 2.14	Simbol ERD	31
Table 3.15	Stop Word Filtering	36
Table 3.16	Stop Word Filtering	37
Table 3.17	Tokenizing	37
Table 3.18	Uji Algoritma Stemming Nazief&Adriani	39
Table 3.19	Contoh Kesalahan Uji Algoritma Stemming Nazief&Adriani	39
Table 3.20	Uji Algoritma Stemming Porter	40
Table 3.21	Contoh Kesalahan Uji Algoritma Stemming Porter	40
Table 4.22	Stop Word Filtering	54
Table 4.23	Case Folding	56
Table 4.24	Tokenizing	56
Table 4.25	Hasil Stemming Algoritma Nazief&Adriani	60
Table 4.26	Hasil Stemming Algoritma Nazief&Adriani	63
Table 4.27	Index Compression Factor	66
Table 4.28	Akurasi Waktu	67
Table 4.29	Proses Stemming Semua Data	68

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	35
Gambar 3.2 Data Flow Diagram (Dfd).....	43
Gambar 3.3 Flowchart Keseluruhan System	44
Gambar 3.4 Flowchart Preprocessing Text	45
Gambar 3.5 Flowchart Filtering	46
Gambar 3.6 Flowchart Filtering Stop Word.....	47
Gambar 3.7 Flowchart Algoritma Stemming Nazief&Adriani	48
Gambar 3.8 Flowchart Algoritma Stemming Porter	49
Gambar 3.9 Struktur Tabel Tb_Katadasar.....	50
Gambar 3.10 Struktur Tabel Tb_Stopword	50
Gambar 3.11 Antarmuka (Interface) Aplikasi	51
Gambar 3.12 Antarmuka (Interface) Aplikasi.....	51
Gambar 4.1 Data Text Stemming.....	52
Gambar 4.2 Script Stop Word Filtering.....	53
Gambar 4.3 Script Case Folding.....	55
Gambar 4.4 Script Stemming Algoritma Nazief &Adriani.....	58
Gambar 4.5 Data Yang Akan Distemming.....	59
Gambar 4.6 Implementasi Stemming Algoritma Nazief&Adriani.....	61
Gambar 4.7 Script Stemming Algoritma Porter	62
Gambar 4.8 Data Yang Akan Distemming.....	63
Gambar 4.9 Implementasi Stemming Algoritma Porter	65
Gambar 4.10 Uji Stemming “Kalimat_Uji2” Algoritma Nazief&Adriani.....	69

INTISARI

Kebutuhan terhadap informasi sangat meningkat. Banyak informasi dibutuhkan untuk dapat diperoleh dengan cepat, dan relevan sesuai kebutuhan. tetapi mengambil informasi yang relevan dan berarti sangat sulit. Berdasarkan hal tersebut dibutuhkan sebuah sistem untuk memenuhi kebutuhan informasi. Stemming merupakan proses untuk memetakan berbagai variasi morfologikal Bahasa Indonesia dari kata menjadi bentuk dasar yang sama, dengan menghilangkan semua imbuhan baik yang terdiri dari awalan, sisipan, akhiran dan kombinasi dari awalan dan akhiran pada kata berimbuhan.

Pada penelitian ini digunakan dua algoritma Nazief dan Adriani dengan Algoritma Stemming Porter, Penelitian ini membandingkan efektifitas dari algoritma stemming Porter dengan Algoritma Nazief dan Adriani untuk stemming Bahasa Indonesia. Untuk membandingkan performa masing-masing algoritma stemming, dibuat program aplikasi web dengan localhost sederhana dengan parameter ketepatan akurasi waktu, word conflation class dan index compression factor. Untuk memudahkan dalam melakukan pengujian maka dilakukan testing model terhadap aplikasi stemming, Selanjutnya akan dilakukan evaluasi terhadap hasil stemming. Evaluasi dilakukan untuk mengetahui proses stemming nilai akurasi dari masing-masing algoritma.

Penulis mengambil judul Analisis Perbandingan Algoritma Stemming Nazief & Adriani dengan Algoritma Stemming Porter. Pembuatan program aplikasi ini bertujuan untuk memudahkan dalam memperoleh informasi dalam bentuk data kata dasar Bahasa Indonesia setelah dilakukan proses stemming sehingga data yang didapat relevan sesuai dengan Kamus Besar Bahasa Indonesia.

Kata kunci : Stemming, Analisis Perbandingan, Algoritma Stemming Nazief & Adriani, Algoritma Stemming Porter.

ABSTRACT

The need for information is greatly increased. A lot of information is needed to be obtained quickly, and relevant as needed. but retrieving relevant and meaningful information is very difficult. Based on this, a system is needed to meet information needs. Stemming is a process to map various morphological variations of the Indonesian language from a word to the same basic form, by eliminating all good imbuhan consisting of prefixes, inserts, suffixes and a combination of prefixes and suffixes in healing words.

In this study used two algorithms Nazief and Adriani with Stemming Porter Algorithm, This study compared the effectiveness of stemming Porter algorithm with Nazief and Adriani algorithm for Stemming Bahasa Indonesia. To compare the performance of each stemming algorithm, a web application program is created with a simple localhost with punctuality parameters, word conflation class and index compression factor. To facilitate in conducting testing, the testing model of stemming application will be conducted next will be evaluated on stemming results. Evaluation is done to find out the process of stemming the accuracy value of each algorithm.

The author takes the title Nazief &Adriani Stemming Algorithm Comparison Analysis with Porter Stemming Algorithm. The creation of this application program aims to facilitate in obtaining information in the form of basic indonesian word data after the stemming process so that the data obtained is relevant in accordance with the Great Dictionary of Bahasa Indonesia.

Keyword : *Stemming, Comparative Analysis, Nazief &Adriani Stemming Algorithm, Porter Stemming Algorithm.*