

# **MODIFIKASI ALGORITMA *STEMMING* SASTRAWI MENGUNAKAN RULE PRECEDENCE DAN PREFIX REMOVAL**

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika



Oleh:

Anwaripasha Akbar

NIM: 09021282025072

Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2024

## LEMBAR PENGESAHAN SKRIPSI

### MODIFIKASI ALGORITMA STEMMING SASTRAWI MENGGUNAKAN RULE PRECEDENCE DAN PREFIX REMOVAL

Oleh:

Anwaripasha Akbar  
NIM: 09021282025072

Mengetahui,  
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.  
NIP. 198005222008121002

Indralaya, 20 Maret 2024

Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.  
NIP. 198410012009121005

## TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 8 Maret 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Anwaripasha Akbar

NIM : 09021282025072

Judul : Modifikasi algoritma *Stemming* Sastrawi menggunakan *Rule Precedence* dan *Prefix Removal*

dan dinyatakan **LULUS**.

1. Ketua

Yunita, M.CS.

NIP 198306062015042002



2. Penguji I

Desty Rodiah, S.Kom., M.T.

NIP 198912212020122011



3. Pembimbing I

Dr. Abdiansah, S.Kom., M.CS.

NIP 198410012009121005



Mengetahui,  
Ketua Jurusan Teknik Informatika



Dr. M. Eachrurrozi, S.Si., M.T.  
NIP 198005222008121002

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Anwaripasha Akbar  
NIM : 09021282025072  
Program Studi : Teknik Informatika  
Judul Skripsi : Modifikasi algoritma *Stemming* Sastrawi menggunakan  
*Rule Precedence* dan *Prefix Removal*

Hasil pengecekan *Software iThenticate/Turnitin*: 16%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak mana pun.



Palembang, 21 Maret 2024



Anwaripasha Akbar  
NIM 09021282025072

## **MOTTO DAN PERSEMBAHAN**

*“To a great mind, nothing is little”*

*– Sir Arthur Conan Doyle*

Kupersembahkan karya tulis ini kepada:

- Allah Subhanahu Wa Ta 'ala dan Rasulullah SAW
- Keluarga saya
- Dosen Pengampu, Pembimbing Akademik dan Skripsi Saya
- Universitas Sriwijaya
- Sahabat dan teman-teman seperjuangan

# MODIFICATION OF SASTRAWI STEMMING ALGORITHM USING PRECEDENCE RULE AND PREFIX REMOVAL

By:  
Anwaripasha Akbar (09021282025072)  
Department of Informatics, Faculty of Computer Science, Sriwijaya University  
Email: pashaakbar52@gmail.com

## ABSTRACT

*Text pre-processing is increasingly important in the era of rapidly growing digital information. One of the important stages in text processing is stemming, which aims to convert words to their basic form by cutting off certain prefixes or suffixes. The Indonesian stemming algorithm that is often used is Sastrawi. To improve the accuracy and efficiency of the stemming process, this research modifies the Sastrawi stemming algorithm using rule precedence and prefix removal. Rule precedence allows better handling of the priority order of removal, while prefix removal modification also aims to improve stemming accuracy. The method was tested through a series of correct stem accuracy tests using datasets from Kompas articles totaling 1666, 1888, 1736, and 1866 words respectively. The experimental results show that the modified Sastrawi stemming outperforms the baseline (Sastrawi Stemmer) with an average stemming accuracy improvement of 0.26%. This research makes an important contribution in the development of a more accurate Indonesian stemming algorithm in Indonesian text processing.*

*Keywords: Stemming, Modified Sastrawi Stemmer, Rule Precedence, Prefix Removal*

Approved,  
Head of Informatics Engineering Department



Dr. M. Fachrurrozi, S.Si., M.T.  
NIP 198005222008121002

Supervisor I,



Dr. Abdiansah, S.Kom., M.Cs  
NIP. 198410012009121005

# MODIFIKASI ALGORITMA *STEMMING* SASTRAWI MENGUNAKAN RULE PRECEDENCE DAN PREFIX REMOVAL

By:

Anwaripasha Akbar (09021282025072)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya



Email: pashaakbar52@gmail.com

## ABSTRAK

Pra-pengolahan teks semakin penting dalam era informasi digital yang berkembang pesat. Salah satu tahapan penting dalam pemrosesan teks adalah *stemming*, yang bertujuan untuk mengubah kata-kata ke bentuk dasarnya dengan memotong awalan atau akhiran tertentu. Algoritma *stemming* Bahasa Indonesia yang sering digunakan adalah Sastrawi. Untuk meningkatkan akurasi dan efisiensi proses *stemming*, penelitian ini melakukan modifikasi algoritma *stemming* Sastrawi menggunakan *rule precedence* dan *prefix removal*. *Rule precedence* memungkinkan penanganan urutan prioritas penghapusan yang lebih baik, sementara modifikasi *prefix removal* juga bertujuan untuk meningkatkan akurasi *stemming*. Metode ini diuji melalui serangkaian pengujian akurasi *correct stem* dengan menggunakan dataset dari artikel Kompas masing-masing berjumlah 1666, 1888, 1736, dan 1866 kata. Hasil eksperimen menunjukkan bahwa modifikasi *stemming* Sastrawi mengungguli *baseline* (Sastrawi Stemmer) dengan peningkatan akurasi *stemming* rata-rata sebesar 0,26%. Penelitian ini memberikan kontribusi penting dalam pengembangan algoritma *stemming* bahasa Indonesia yang lebih akurat dalam pemrosesan teks Bahasa Indonesia.

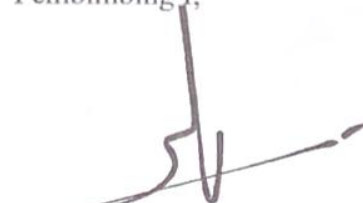
Kata Kunci: *Stemming*, Sastrawi Stemmer, *Rule Precedence*, *Prefix Removal*

Mengetahui,  
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.  
NIP. 198005222008121002

Pembimbing I,



Dr. Abdiansah, S.Kom., M.Cs  
NIP. 198410012009121005

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT, Tuhan Semesta Alam atas berkat, rahmat, rahim dan karunia-Nya yang telah diberikan kepada penulis sehingga Tugas Akhir berjudul “Modifikasi algoritma *Stemming* Sastrawi menggunakan *Rule Precedence* dan *Prefix Removal*” dapat disusun dengan baik sebagai syarat dalam menyelesaikan studi Strata-1 program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Sriwijaya. Pada kesempatan ini penulis ingin mengucapkan ucapan terima kasih banyak kepada semua pihak yang memberikan dukungan, motivasi dan bimbingan selama penyusunan Tugas Akhir dan penelitian. Secara khusus ucapan terima kasih ini ditujukan kepada:

1. Allah Subhanallahu Wa Ta’ala atas segala berkah, rahmat, rahim dan karunia-Nya.
2. Keluarga saya yang selalu memberikan dukungan positif.
3. Prof. DR. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Dr. M. Fachrurrozi, S.Si., M.T selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku Dosen Pembimbing saya yang senantiasa membagikan ilmu, membimbing, memberi arahan dan memotivasi saya untuk menyelesaikan Tugas Akhir ini dengan baik.



6. Ibu Nabila Rizky Oktadini, M.T. selaku Dosen Pembimbing Akademik yang senantiasa memberikan saran, arahan dan bimbingan kepada saya selama masa perkuliahan.
7. Seluruh Dosen Fakultas Ilmu Komputer, khususnya Dosen Jurusan Teknik Informatika, yang telah banyak memberikan bekal, arahan, dan saran selama masa perkuliahan.
8. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.
9. Teman – temanku yang telah membantu dan memberikan dukungan secara langsung maupun tidak langsung dalam proses pengerjaan Tugas Akhir ini.

Penulis secara penuh menyadari akan kekurangan dalam penyusunan Tugas Akhir ini. Kekurangan ini semata – mata adalah karena keterbatasan pengetahuan juga pengalaman penulis. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk menyempurnakan Tugas akhir ini sehingga dapat membawa manfaat lebih banyak dan luas. Akhir kata, semoga Tugas Akhir ini bermanfaat bagi kita semua.

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI .....	ii
HALAMAN PERNYATAAN .....	iii
MOTTO DAN PERSEMBAHAN .....	iv
ABSTRACT.....	v
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
BAB I PENDAHULUAN.....	I-1
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang .....	I-1
1.3    Rumusan Masalah .....	I-3
1.4    Tujuan Penelitian.....	I-3
1.5    Manfaat Penelitian.....	I-3
1.6    Batasan Masalah.....	I-4
1.7    Sistematika Penulisan.....	I-4
1.8    Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR.....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori .....	II-1
2.2.1    Imbuhan Bahasa Indonesia .....	II-1
2.2.2 <i>Stemming</i> .....	II-3
2.2.2    Tokenisasi.....	II-5
2.2.3    Nazief Adriani <i>Stemmer</i> .....	II-6
2.2.4 <i>Confix Stripping Stemmer</i> .....	II-7
2.2.5 <i>Enhanced Confix Stripping Stemmer</i> .....	II-11
2.2.6    Rational Unified Process.....	II-13
2.3    Penelitian Lain yang Relevan.....	II-15

2.4	Kesimpulan.....	II-16
BAB III METODE PENELITIAN.....		III-1
3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data .....	III-1
3.3	Tahapan Penelitian.....	III-2
3.3.1	Kerangka Kerja .....	III-2
3.3.2	Kriteria Pengujian .....	III-5
3.3.3	Format Data Pengujian.....	III-6
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian .....	III-7
3.3.5	Pengujian Penelitian.....	III-7
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-8
3.4	Metode Pengembangan Perangkat Lunak .....	III-9
3.4.1	Fase Insepsi .....	III-9
3.4.2	Fase Elaborasi .....	III-9
3.4.3	Fase Konstruksi.....	III-9
3.4.4	Fase Transisi.....	III-10
3.5	Kesimpulan.....	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK .....		IV-1
4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-1
4.2.2	Kebutuhan Sistem .....	IV-2
4.2.3	Analisis dan Desain.....	IV-3
4.3	Fase Elaborasi.....	IV-13
4.3.1	Pemodelan Bisnis .....	IV-14
4.3.2	Kebutuhan Sistem .....	IV-16
4.3.3	Diagram.....	IV-17
4.4	Fase Konstruksi .....	IV-21
4.4.1	Kebutuhan .....	IV-21
4.4.2	Implementasi.....	IV-22
4.5	Fase Transisi.....	IV-25
4.5.1	Pemodelan Sistem .....	IV-25

4.5.2	Kebutuhan .....	IV-25
4.5.3	Analisis dan Perancangan .....	IV-26
4.5.4	Implementasi .....	IV-28
4.6	Kesimpulan.....	IV-30
BAB V HASIL DAN ANALISIS PENELITIAN .....		V-1
5.1	Pendahuluan .....	V-1
5.2	Hasil Analisis Pustaka Sastrawi .....	V-1
5.3.1	Data Kata Dasar Bahasa Indonesia .....	V-1
5.3.2	Analisis Tambahan Aturan Algoritma Sastrawi .....	V-1
5.3.3	Analisis Kesalahan <i>Stemming</i> Sastrawi .....	V-2
5.3	Usulan Skema Perbaikan .....	V-6
5.4	Hasil Program.....	V-8
5.5	Hasil Pengujian.....	V-12
5.3.1	Konfigurasi Pengujian.....	V-12
5.3.1	Hasil Analisis Pengujian .....	V-13
5.3.2	Hasil Analisis Kesalahan <i>Stemming</i> .....	V-15
5.6	Analisis Hasil Penelitian .....	V-19
5.6.1	Analisis Hasil Kesalahan <i>Stemming</i> .....	V-19
5.7	Kesimpulan.....	V-23
BAB VI KESIMPULAN DAN SARAN .....		VI-1
6.1	Pendahuluan .....	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA .....		
LAMPIRAN.....		

## DAFTAR TABEL

<b>Tabel II- 1.</b> Aturan pemenggalan prefiks Nazief Adriani <i>Stemmer</i> .....	II-9
<b>Tabel II- 2.</b> Modifikasi tambahan aturan oleh algoritma CSS .....	II-11
<b>Tabel II- 3.</b> Modifikasi tambahan aturan oleh algoritma ECS .....	II-11
<b>Tabel III- 1.</b> Rancangan tabel hasil <i>stemming</i> .....	III-6
<b>Tabel III- 2.</b> Rancangan Tabel Analisis Hasil Pengujian .....	III-8
<b>Tabel IV- 1.</b> Tabel Kebutuhan Fungsional.....	IV-2
<b>Tabel IV- 2.</b> Tabel Kebutuhan Non-Fungsional .....	IV-3
<b>Tabel IV- 3.</b> Contoh data teks artikel.....	IV-5
<b>Tabel IV- 4.</b> Hasil Normalisasi data teks artikel.....	IV-5
<b>Tabel IV- 5.</b> Hasil Tokenisasi data teks artikel .....	IV-6
<b>Tabel IV- 6.</b> Hasil <i>Stemming</i> MSS data teks artikel .....	IV-7
<b>Tabel IV- 7.</b> Tabel Definisi Aktor .....	IV-10
<b>Tabel IV- 8.</b> Tabel Definisi <i>Use Case</i> .....	IV-10
<b>Tabel IV- 9.</b> Skenario Use Case <i>Stemming</i> menggunakan MSS .....	IV-11
<b>Tabel IV- 10.</b> Skenario Use Case Mengkalkulasi <i>correct stem Stemmer</i> .....	IV-12
<b>Tabel IV- 11.</b> Keterangan Implementasi Kelas.....	IV-23
<b>Tabel IV- 12.</b> Rencana Pengujian <i>Stemming</i> menggunakan MSS.....	IV-27
<b>Tabel IV- 13.</b> Rencana Pengujian Mengkalkulasi <i>correct stem Stemmer</i> .....	IV-27
<b>Tabel IV- 14.</b> Pengujian Use Case <i>Stemming</i> menggunakan MSS.....	IV-28
<b>Tabel IV- 15.</b> Pengujian Use Case Mengkalkulasi <i>correct stem Stemmer</i> .....	IV-29
<b>Tabel V- 1.</b> Modifikasi Tambahan Aturan oleh Algoritma Sastrawi <i>Stemmer</i> .....	V-2
<b>Tabel V- 2.</b> Kesalahan <i>Stemming</i> Sastrawi .....	V-2
<b>Tabel V- 3.</b> Tabel Hasil Analisis Pengujian Data Artikel 1 .....	V-13
<b>Tabel V- 4.</b> Tabel Hasil Analisis Pengujian Data Artikel 2.....	V-13
<b>Tabel V- 5.</b> Tabel Hasil Analisis Pengujian Data Artikel 3.....	V-14
<b>Tabel V- 6.</b> Tabel Hasil Analisis Pengujian Data Artikel 4.....	V-14
<b>Tabel V- 7.</b> Tabel Hasil Analisis Hasil <i>Stem</i> Data Artikel 1 .....	V-15
<b>Tabel V- 8.</b> Tabel Hasil Analisis Hasil <i>Stem</i> Data Artikel 2 .....	V-16
<b>Tabel V- 9.</b> Tabel Hasil Analisis Hasil <i>Stem</i> Data Artikel 3 .....	V-17
<b>Tabel V- 10.</b> Tabel Hasil Analisis Hasil <i>Stem</i> Data Artikel 4 .....	V-18

## DAFTAR GAMBAR

<b>Gambar II- 1.</b> Fase-fase dari RUP .....	II-14
<b>Gambar III- 1.</b> Diagram Tahapan Penelitian.....	III-2
<b>Gambar III- 2.</b> Kerangka Kerja Penelitian .....	III-3
<b>Gambar III- 3.</b> Diagram Proses Modifikasi Sastrawi <i>Stemmer</i> .....	III-5
<b>Gambar IV- 1.</b> Diagram <i>Use Case</i> .....	IV-9
<b>Gambar IV- 2.</b> Rancangan Antarmuka <i>Modified Sastrawi Stemmer</i> .....	IV-15
<b>Gambar IV- 3.</b> Rancangan Antarmuka Analisis <i>Stemming</i> .....	IV-15
<b>Gambar IV- 4.</b> Diagram Aktivitas <i>Stemming</i> Menggunakan MSS .....	IV-18
<b>Gambar IV- 5.</b> Diagram Aktivitas Mengkalkulasi <i>correct stem Stemmer</i> .....	IV-19
<b>Gambar IV- 6.</b> <i>Sequence Diagram Stemming</i> menggunakan MSS .....	IV-20
<b>Gambar IV- 7.</b> <i>Sequence Diagram</i> Mengkalkulasi <i>correct stem Stemmer</i> ....	IV-21
<b>Gambar IV- 8.</b> Diagram Kelas.....	IV-22
<b>Gambar IV- 9.</b> Tampilan Antarmuka <i>Modified Sastrawi Stemmer</i> .....	IV-24
<b>Gambar IV- 10.</b> Tampilan Antarmuka Analisis Hasil <i>Stemming</i> .....	IV-25
<b>Gambar V- 1.</b> Tampilan <i>Input Data</i> Teks Artikel Menggunakan <i>Text Area</i> .....	V-8
<b>Gambar V- 2.</b> Tampilan <i>Input Data</i> Teks Artikel Menggunakan <i>Upload File</i> ....	V-9
<b>Gambar V- 3.</b> Tampilan Teks Setelah Dilakukan <i>Stemming</i> .....	V-9
<b>Gambar V- 4.</b> Tampilan Halaman Analisis <i>Stemming</i> .....	V-10
<b>Gambar V- 5.</b> Tampilan <i>Input Data</i> Teks Artikel (Analisis) .....	V-10
<b>Gambar V- 6.</b> Tampilan <i>Input Data</i> Kata Dasar Validasi Artikel (Analisis).....	V-11
<b>Gambar V- 7.</b> Hasil Analisis Perbandingan MSS Dengan <i>Stemmer</i> Lainnya... V-11	
<b>Gambar V- 8.</b> Perbandingan <i>Correct Stem</i> dan <i>Incorrect Stem</i> Artikel 1 .....	V-20
<b>Gambar V- 9.</b> Perbandingan <i>Correct Stem</i> dan <i>Incorrect Stem</i> Artikel 2 .....	V-21
<b>Gambar V- 10.</b> Perbandingan <i>Correct Stem</i> dan <i>Incorrect Stem</i> Artikel 3 .....	V-22
<b>Gambar V- 11.</b> Perbandingan <i>Correct Stem</i> dan <i>Incorrect Stem</i> Artikel 4 .....	V-23

## DAFTAR LAMPIRAN

1. Kode Program
2. Data Teks Artikel
3. Hasil Normalisasi Data Teks Artikel
4. Hasil Tokenisasi Data Teks Artikel
5. Hasil *Stemming* MSS Data Teks Artikel

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab pendahuluan ini menjelaskan dasar-dasar pemikiran yang menjadi dasar dari rencana penelitian. Dasar-dasar pemikiran ini mencakup latar belakang penelitian, perumusan masalah atau permasalahan penelitian, tujuan penelitian, dan manfaat penelitian.

### **1.2 Latar Belakang**

Dalam era informasi digital, kecepatan dan akurasi dalam menavigasi serta memahami konten menjadi krusial. Perkembangan teknologi informasi dan komunikasi telah meningkatkan volume data yang tersedia secara *online*, termasuk dokumen-dokumen ilmiah dan sumber informasi lainnya. Dalam konteks ini, kemampuan untuk mengidentifikasi dan mengekstrak informasi yang relevan, seperti kata kunci, dari kumpulan data besar menjadi sangat penting (Brotosaputro *et al.*, 2022).

Teknik *stemming*, dalam pemrosesan bahasa alami, memainkan peran vital dalam kedua bidang ini. Dengan mengurangi kata-kata ke bentuk dasar atau '*root*', *stemming* memudahkan pencocokan antara istilah pencarian dan dokumen, serta membantu dalam proses ekstraksi kata kunci yang lebih efisien (Novitasari, 2017). Dengan mengeliminasi variasi morfologi, *stemming* meningkatkan relevansi dan fokus dalam pencarian informasi, serta efisiensi dalam mengidentifikasi kata kunci



yang penting. Dengan demikian, proses ini tidak hanya meningkatkan kualitas hasil pencarian tetapi juga mempercepat akses ke informasi yang dibutuhkan.

Khususnya dalam Bahasa Indonesia, struktur morfologi yang unik dengan beragam prefiks, sufiks, infiks, dan konfiks menjadikan prosesnya lebih rumit. Ini berarti ada lebih banyak aspek yang harus dipertimbangkan saat melakukan *stemming*, sehingga memerlukan pendekatan yang lebih teliti (Wahyudi *et al.*, 2017).

Algoritma *Stemming* Sastrawi telah menjadi algoritma yang populer untuk mengatasi kompleksitas ini. Namun, masih peluang dilakukan modifikasi lebih lanjut. Penelitian ini fokus pada modifikasi algoritma ini dengan mengintegrasikan *rule precedence* dan *prefix removal* yang baru. Tujuan utama dari penambahan ini adalah untuk meningkatkan akurasi dan efisiensi dalam mengidentifikasi bentuk dasar kata. *Rule precedence*, yang menetapkan urutan penghapusan imbuhan, diperkirakan dapat secara signifikan mempengaruhi hasil *stemming*. Modifikasi *prefix removal* menjadi kunci dalam upaya mencegah kesalahan *stem* pada algoritma *stemming* Sastrawi. Dengan menyesuaikan urutan ini, diharapkan algoritma akan menjadi lebih adaptif dalam menavigasi struktur morfologi Bahasa Indonesia yang kompleks.

Menurut beberapa penelitian terdahulu, telah ada Algoritma *stemming* Sastrawi merupakan algoritma *stemming* yang populer dipakai di *stemming* Bahasa Indonesia. Meskipun telah banyak penelitian sebelumnya yang mengeksplorasi penggunaan algoritma *stemming* Sastrawi, namun masih terdapat sedikit penelitian yang bertujuan untuk meningkatkan akurasi dari algoritma Sastrawi itu sendiri.

Pemilihan untuk memfokuskan pada modifikasi *rule precedence* dan *prefix removal* ini sangat penting. Modifikasi tersebut diharapkan tidak hanya untuk mengurangi kesalahan *stemming*, tetapi juga meningkatkan akurasi keseluruhan algoritma. Dengan memanfaatkan algoritma Sastrawi sebagai dan memodifikasi *rule precedence* dan *prefix removal* dalam algoritma *stemming* Sastrawi, peneliti bertujuan untuk memungkinkan pemrosesan teks dengan lebih akurat.

### 1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana memodifikasi algoritma *Stemming* Sastrawi?
2. Bagaimana kinerja algoritma *Stemming* Sastrawi setelah dimodifikasi?

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Menghasilkan algoritma *stemming* Sastrawi yang dimodifikasi dalam pengolahan Bahasa Indonesia.
2. Mengetahui kinerja algoritma *stemming* Sastrawi yang telah dimodifikasi.

### 1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini sebagai berikut:

1. Berkontribusi pada pengembangan NLP untuk Bahasa Indonesia melalui optimasi algoritma *stemming* Sastrawi.

2. Temuan dari penelitian ini bisa menjadi acuan bagi studi-studi mendatang yang terkait dengan bidang *stemming* Bahasa Indonesia.

## 1.6 Batasan Masalah

Batasan masalah dalam penelitian ini sebagai berikut:

1. Objek penelitian merupakan teks berbahasa Indonesia.
2. Data teks artikel untuk penelitian secara manual dari artikel dari Kompas dan data kamus kata dasar diambil dari kateglo.
3. Perbandingan algoritma hanya mencakup algoritma *stemming* Sastrawi yang sudah dimodifikasi, Sastrawi, *Confix Stripping Stemmer*, dan *Enhanced Confix Stripping Stemmer*.

## 1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

### **BAB I. PENDAHULUAN**

Bab ini menjelaskan tentang latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

### **BAB II. KAJIAN LITERATUR**

Bab ini menjelaskan tentang dasar-dasar teori yang digunakan dalam penelitian, seperti Imbuhan Bahasa Indonesia, *Stemmer*, *Natural Language Processing*, *Nazief Adriani Stemmer*, *Confix Stripping Stemmer*, *Enhanced Confix Stripping Stemmer*, dan Tokenisasi.

### **BAB III. METODOLOGI PENELITIAN**

Bab ini menjelaskan langkah-langkah yang akan dilakukan dalam penelitian ini. Setiap rencana fase penelitian dirinci dengan mengacu pada kerangka kerja. Di akhir bab ini, perancangan manajemen proyek disertakan saat melakukan penelitian.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini akan dijelaskan proses pengembangan perangkat lunak modifikasi algoritma *stemming* Sastrawi menggunakan penambahan aturan baru dalam *rule precedence* dan *prefix removal*. Pengembangan perangkat lunak ini akan menggunakan metode *Rational Unified Process*.

### **BAB V. HASIL DAN ANALISA PENELITIAN**

Pada bab ini akan dijelaskan mengenai hasil dari penelitian dalam mengembalikan suatu kata menjadi kata dasar melalui modifikasi algoritma *stemming* Sastrawi menggunakan penambahan aturan baru dalam *rule precedence* dan *prefix removal*

### **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari penelitian yang telah dilakukan sebelumnya dan juga berisi saran yang dapat dilakukan pada penelitian selanjutnya yang berhubungan dengan modifikasi algoritma *stemming* Sastrawi menggunakan penambahan aturan baru dalam *rule precedence* dan *prefix removal*.

## **1.8 Kesimpulan**

Dalam bab ini, telah dibahas secara komprehensif poin-poin penting yang menjadi dasar dalam penelitian, seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan struktur penulisan.

## DAFTAR PUSTAKA

- Adriani, M., Asian, J., Nazief, B., Williams, H. E., & Tahaghoghi, S. M. M. (2007). Stemming Indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing*, 6(4), 13. <https://doi.org/10.1145/1316457.1316459>
- Alfando, A., & Hayami, R. (2023). Klasifikasi Teks Berita Berbahasa Indonesia Menggunakan Machine Learning Dan Deep Learning. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 681–686. <https://doi.org/10.36040/JATI.V7I1.6486>
- Arifin, A. Z., Mahendra, I. P. A. K., & Ciptaningtyas, H. T. (2009). Enhanced confix stripping stemmer and ants algorithm for classifying news document in Indonesian language. *In The International Conference on Information & Communication Technology and Systems* (Vol. 5, pp. 149-158).
- Bisma, R., Putra, S., Utami, E., & Raharjo, S. (2018). Optimalisasi Stemming Kata Berimbuhan Tidak Baku Pada Bahasa Indonesia Dengan Levenshtein Distance. *Jurnal Informatika: Jurnal Pengembangan IT*, 3(2), 200–205.
- Brotosaputro, G., Windihastuty, W., Anugrah Mutiarawan, R., Fakultas, J., & Informasi, T. (2022). Penentuan Hoax pada Artikel Politik Berbahasa Indonesia di Sosial Media dengan Similarity Jaccard dan Algoritma Stemming. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(1), 79–86. <https://doi.org/10.32736/sisfokom.v11i1.1358>

- Chaer, A. (2011). Tata bahasa praktis bahasa Indonesia.. *Rineka Cipta Rev. ed.*
- Hudin, M. S., Fauzi, M. A., & Adinugroho, S. (2018). Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 5518–5524.
- Kruchten, P. (2004). *The Rational unified Process An Introduction (Third).*
- Lelywiary, C. J. L., Widowati, S., & Muslim, K. L. (2019). Deteksi Pola Ambiguitas Struktural pada Spesifikasi Kebutuhan Perangkat Lunak menggunakan Pemrosesan Bahasa Alami. *Indonesia Journal on Computing (Indo-JC)*, 4(3), 51–64. <https://doi.org/10.34818/INDOJC.2019.4.3.355>
- Magriyanti, A. A. (2018). Analisis Pengembangan Algoritma Porter Stemming Dalam Bahasa Indonesia. <https://doi.org/10.31227/OSF.IO/7GE4V>
- Mantik, J., Rosidin, M., Fauzan Gustafi, M., Pratiwi, S. A., Teknologi, F., Kesehatan, D. I., & Informasi, S. (2023). Optimizing Nazief Adriani's Stemmer Algorithm In Detecting Indonesian Word Errors Using Sastrawi. *Jurnal Mantik*, 7(3), 1733–1742. <https://doi.org/10.35335/MANTIK.V7I3.4020>
- Fradana, A. N. (2018). Buku Ajar Morfologi Bahasa. *Umsida Press*. <https://doi.org/10.21070/2018/978-602-5914-31-7>
- Muchtar, M. A., Jaya, I., Nababan, M., Andayani, U., Siregar, L. N., Nababan, E. B., & Sitompul, O. S. (2019). Separation of Basic Words in Angkola Batak

Text Documents using Enhanced Confix Stripping Stemmer Case: Mandailing Ethnic. *IOP Conference Series: Materials Science and Engineering*, 648(1), 012024. <https://doi.org/10.1088/1757-899X/648/1/012024>

Novitasari, D. (2017). Perbandingan Algoritma Stemming Porter dengan Arifin Setiono untuk Menentukan Tingkat Ketepatan Kata Dasar. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 1(2), 120–129.

Perwitasari, R., Afawani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) Dalam Pengembangan Sistem Informasi Medical Check Up Pada Citra Medical Centre. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, 2(1), 76–88. <https://doi.org/10.29303/JTIKA.V2I1.85>

Rofiqi, M. A., Fauzan, A. C., Agustin, A. P., Saputra, A. A., & Fahma, H. D. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 58–64. <https://doi.org/10.28926/ILKOMNIKA.V1I2.18>

Simarankir, M. S. H. (2017). Studi Perbandingan Algoritma - Algoritma Stemming Untuk Dokumen Teks Bahasa Indonesia. *Jurnal Inkofar*, 1(1), 2581–2920.

Simpem, I. (2021). *Morfologi: Kajian proses pembentukan kata*. Bumi Aksara



- Song, X., Salcianu, A., Song, Y., Dopson, D., & Zhou, D. (2020). Fast WordPiece Tokenization. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2089–2103.  
<https://doi.org/10.18653/v1/2021.emnlp-main.160>
- Sowmya V. B., Majumder, B., Gupta, A. , & Surana, H. (2020). *Practical Natural Language Processing: A Pragmatic Approach to Processing and Analyzing Language Data*. O'Reilly.
- Sugiyamta. (2015). Sistem Deteksi Kemiripan Dokumen Dengan Algoritma Cosine Similarity Dan Single Pass Clustering. *Dinamika Informatika*, 7(2).
- Wahyudi, D., Susyanto, & T., Nugroho, D. (2017). Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah SINUS*, 15(2).  
<https://doi.org/10.30646/SINUS.V15I2.305>