

**PEMROSESAN PARALEL UNTUK *STEMMING* DAN
POS-TAGGING PADA TEKS BAHASA INDONESIA**

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



Oleh:

Muhammad Wahyu Rahmadani
NIM : 09021181823166

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022**

LEMBAR PENGESAHAN SKRIPSI


PEMROSESAN PARALEL UNTUK *STEMMING* DAN *POS-TAGGING* PADA TEKS BAHASA INDONESIA

Oleh :

Muhammad Wahyu Rahmadani
NIM : 09021181823166

Indralaya, 20 Desember 2022

Pembimbing I



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

Pembimbing II,



Mastura Diana Marieska, M.T
NIP. 198603212018032001

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Senin tanggal 12 Desember 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Muhammad Wahyu Rahmadani
NIM : 09021181823166
Judul : Pemrosesan Paralel untuk *Stemming* dan *POS-Tagging* pada Teks Bahasa Indonesia

1. Ketua Penguji

Rizki Kurniati, M.T.
NIP. 199107122019032016



2. Pembimbing I

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



3. Pembimbing II

Mastura Diana Marieska, M.T.
NIP. 198603212018032001



4. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001



5. Penguji II

M. Qurhanul Rizqie, M.T., Ph.D.
NIDN. 0203128701



Mengetahui,
Ketua Jurusan Teknik Informatika



Alva Syahmin Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertandatangan dibawah ini:

Nama : Muhammad Wahyu Rahmadani

NIM : 09021181823166

Jurusan : Teknik Informatika

Judul Skripsi : Pemrosesan Paralel untuk *Stemming* dan *POS-Tagging* pada Teks Bahasa Indonesia


Hasil pengecekan *Software iThenticate/Turnitin* : 6 %

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 oleh siapapun.



Indralaya, 20 Desember 2022



Muhammad Wahyu Rahmadani
NIM. 09021181823166

MOTTO DAN PERSEMBAHAN

Motto:

“There is nothing so practical as a good theory,

There is nothing so theoretical as a good practice.” - Kurt Lewin-

Persembahan Karya Tulis ini Kepada:

- Ibuk & Alm. Ayah
- Ibu Dian Palupi Rini, M.Kom, Ph.D., Dosen Pembimbing Akademik
- Bapak Dr. Abdiansah, S.Kom., M.CS, Dosen Pembimbing Skripsi Penulis
- Ibu Mastura Diana Marieska, M.T, Dosen Pembimbing Skripsi Penulis
- Saudara-saudaraku
- Kawan Seperjuangan Masa Perkuliahan

ABSTRACT

Stemming and *POS-Tagging* is part of the pre-processing of raw text data in the natural language processing field which aims to produce more structured data and as an initial step that greatly affects processing performance before being processed at a further stage. In its application for Indonesian text, the efficiency level of process performance for these two stages is still low, especially for large data sizes. Parallel processing method using the *python multiprocessing module* was applied in this study to see the reduction in processing time for the *Stemming* and *POS-Tagging* process and also to observe the impact of implementing this parallel processing method on the devices used. Results showed that the highest reduction was 78.26% for the *POS-Tagging* process using test data size range of 10 MB – 70 MB and 63.28% for the *Stemming* process using test data size range of 7 MB – 25 MB. Processor allocation in parallel processing and data size affect device performance in terms of increasing device temperature and memory consumption.

Keywords: parallel processing, *Stemming*, *POS-Tagging*. Indonesian text

ABSTRAK

Stemming dan *POS-Tagging* merupakan bagian dari tahap pra-pengolahan data mentah teks pada bidang pemrosesan bahasa alami yang bertujuan menghasilkan data yang lebih terstruktur dan sebagai langkah awal yang sangat mempengaruhi performa pemrosesan sebelum diproses pada tahap lebih lanjut. Dalam pengaplikasiannya pada teks berbahasa Indonesia, tingkat efisiensi waktu kinerja proses untuk kedua tahap ini masih rendah terutama untuk ukuran data yang besar. Metode pemrosesan paralel menggunakan modul *multiprocessing python* diterapkan dalam penelitian ini untuk melihat tingkat reduksi waktu pengolahan pada tahap *Stemming* dan *POS-Tagging* yang dihasilkan serta melakukan pengamatan terhadap dampak penerapan metode pemrosesan paralel pada perangkat yang digunakan. Hasil penelitian menunjukkan reduksi tertinggi 78,26 % untuk tahap *POS-Tagging* pada rentang ukuran data uji 10 MB – 70 MB dan 63,28% untuk tahap *Stemming* pada rentang ukuran data uji 7 MB – 25 MB. Alokasi prosesor pada pemrosesan paralel dan ukuran data mempengaruhi performa kinerja perangkat dari sisi peningkatan suhu perangkat dan konsumsi memori.

Kata Kunci: Pemrosesan Paralel, *Stemming*, *POS-Tagging*, Teks Bahasa Indonesia

KATA PENGANTAR

Segala puji hanya milik Allah swt. Tuhan semesta alam, Tuhan yang maha Agung, maha Pemberi Kecukupan, maha Pemberi Petunjuk, maha Mengetahui lagi Maha Penyayang. Syukur yang tak terhingga karena hanya atas ridho dan pertolongan-Nya penulis dapat menyelesaikan penulisan tugas akhir dengan judul “Pemrosesan Paralel untuk *Stemming* dan *POS-Tagging* pada Teks Bahasa Indonesia” ini.

Pada kesempatan ini, penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bantuan, motivasi, pembelajaran, pengalaman, juga ilmu yang sangat berharga kepada penulis selama berproses pada jenjang pendidikan Strata-I ini. Secara khusus, penulis ucapkan terima kasih ini kepada:

1. Allah *subhanahu wa ta'ala*...
2. Rasulullah SAW. (*Allahumma Sholli 'Ala Sayyidina Muhammad*)
3. Ibuk & Almarhum Ayah, surga pertamaku,
juga Saudara-saudaraku, *support system* terbaik, rumah tempat kembali.
4. Bapak Jaidan Jauhari, M.T., Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Alvi Syahrini Utami, M.Kom., Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Ibu Dian Palupi Rini, M.Kom, Ph.D., Dosen Pembimbing Akademik yang senantiasa memberikan arahan selama masa perkuliahan.

7. Bapak Abdiansah, S.Kom., M.Cs. dan Ibu Mastura Diana Marieska, M.T., Dosen Pembimbing Skripsi penulis yang telah banyak memberikan ilmu, motivasi, pembelajaran, arahan dan didikan yang sangat berharga di masa penyusunan tugas akhir ini.
8. Bapak-Ibu dosen pendidik penulis selama masa perkuliahan, semoga menjadikan amal jariyah untuk Bapak-Ibu atas ilmu yang telah diajarkan.
9. Sahabat sekaligus saudaraku selama masa perkuliahan, Rama. Juga sahabat-sahabat terbaikku yang lain sahabat seperjuangan, semoga kalian tetap dalam lindungan Allah swt.
10. *Staff* Fasilkom yang senantiasa mewadahi dan memberikan bantuan, kak irawan admin perpus, kak cokro admin lab, kak ricy admin jurusan, petugas kebersihan di ilkom yang selalu ramah buat saling menyapa dan seluruh *staff* yang membantu penulis selama belajar di fasilkom unsri layo.

Penulis menyadari kekurangan-kekurangan yang masih perlu dikoreksi dalam penyusunan Tugas Akhir ini yang semata-mata karena keterbatasan ilmu pengetahuan dan pengalaman penulis. Kritik dan saran diharapkan untuk pengembangan lebih lanjut pada topik Tugas Akhir ini agar lebih banyak membawa kebermanfaatan. Akhir kata, semoga karya tulis ini bermanfaat dan membawa kebaikan untuk pembaca.

Indralaya, Oktober 2022

Muhammad Wahyu Rahmadani

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
ABSTRACT.....	ii
ABSTRAK.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiv
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.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-4
1.7 Sistematika Penulisan.....	I-5
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 Pemrosesan Paralel.....	II-1
2.2.2 <i>Python Multiprocessing Module</i>	II-4

2.2.2 <i>Python Time Module</i>	II-4
2.2.3 <i>Stemming</i>	II-5
2.2.4 <i>POS-Tagging</i>	II-9
2.2.5 Tata Bahasa Indonesia.....	II-12
2.2.6 <i>Rational Unified Process</i>	II-14
2.2.7 Pengujian.....	II-17
2.3 Penelitian Lain yang Relevan.....	II-19
2.3.1 Yusliani, Primartha dan Diana (2019).....	II-19
2.3.2 Nurwidyantoro dan Winarko (2012).....	II-20
2.4 Kesimpulan.....	II-21
BAB III METODE PENELITIAN	III-1
3.1 Pendahuluan.....	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis dan Sumber Data.....	III-1
3.2.2 Metode Pengumpulan Data.....	III-2
3.3 Tahapan Penelitian.....	III-3
3.3.1 Pengujian Penelitian Menggunakan Perangkat Lunak.....	III-6
3.3.2 Kriteria Pengujian.....	III-8
3.3.3 Format Data Pengujian.....	III-8
3.3.4 Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-10
3.3.5 Pengujian Penelitian.....	III-11
3.3.6 Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-12
3.4 Metode Pengembangan Perangkat Lunak.....	III-13

3.4.1 Fase Insepsi.....	III-13
3.4.2 Fase Elaborasi.....	III-13
3.4.3 Fase Konstruksi.....	III-13
3.4.4 Fase Transisi.....	III-14
3.5 Kesimpulan.....	III-14
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-4
4.2.3.1 Analisis dan Kebutuhan Perangkat Lunak.....	IV-4
4.2.3.2 Analisis Data.....	IV-5
4.2.3.3 Analisis Metode Pemrosesan Serial dan Pemrosesan Paralel.....	IV-5
4.2.3.4 Desain Perangkat Lunak.....	IV-6
1. Diagram <i>Use Case</i>	IV-7
2. Definisi Aktor.....	IV-8
3. Definisi <i>Use Case</i>	IV-8
4. Skenario <i>Use Case</i>	IV-9
4.3 Fase Elaborasi.....	IV-13
4.3.1 Pemodelan Bisnis.....	IV-13
4.3.2 Perancangan Data.....	IV-14

4.3.3	Kebutuhan Sistem.....	IV-14
4.3.4	Diagram.....	IV-15
4.3.4.1	<i>Activity Diagram</i>	IV-15
4.3.4.2	<i>Sequence Diagram</i>	IV-17
4.3.5	Perancangan Antarmuka.....	IV-20
4.4	Fase Konstruksi.....	IV-20
4.4.1	Kebutuhan Sistem.....	IV-20
4.4.2	Diagram Kelas.....	IV-21
4.4.3	Implementasi.....	IV-21
4.4.3.1	Implementasi Kelas.....	IV-21
4.4.3.2	Implementasi Antarmuka.....	IV-23
4.5	Fase Transisi.....	IV-23
4.5.1	Pemodelan Bisnis.....	IV-23
4.5.2	Kebutuhan.....	IV-24
4.5.3	Analisis dan Perancangan.....	IV-25
4.5.4	Implementasi.....	IV-26
4.5.5	Kesimpulan.....	IV-27
BAB V	HASIL DAN PEMBAHASAN.....	V-1
5.1	Pendahuluan.....	V-1
5.2	Hasil Penelitian.....	V-1
5.2.1	Konfigurasi Pengujian.....	V-1
	Data Hasil Pengujian Tahap I-1.....	V-2
	Data Hasil Pengujian Tahap I-2.....	V-3

Data Hasil Pengujian Tahap I-3.....	V-5
Data Hasil Pengujian Tahap I-4.....	V-7
Data Hasil Pengujian Tahap II-1.....	V-9
Data Hasil Pengujian Tahap II-2.....	V-10
Data Hasil Pengujian Tahap II-3.....	V-12
Data Hasil Pengujian Tahap II-4.....	V-14
5.3 Analisis Hasil Penelitian.....	V-17
5.3.1 Analisis Hasil Pengujian Performansi Waktu Proses.....	V-17
5.4 Kesimpulan.....	V-21
BAB VI KESIMPULAN DAN SARAN.....	VI-1
6.1 Kesimpulan.....	VI-2
6.2 Saran.....	VI-2

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel II-1	: <i>Tagset</i> untuk Bahasa Indonesia.....	II-11
Tabel II-2	: Penguraian kalimat berdasarkan jenis Pengklasifikasi.....	II-15
Tabel III-1	: Pembagian Ukuran Data Uji Proses <i>Stemming</i>	III-2
Tabel III-2	: Pembagian Ukuran Data Uji Proses <i>POS-Tagging</i>	III-2
Tabel III-3	: Format data hasil pengujian proses <i>Stemming</i>	III-8
Tabel III-4	: Format data hasil pengujian proses <i>POS-Tagging</i>	III-9
Tabel IV-1	: Kebutuhan Fungsional Perangkat Lunak Pengujian.....	IV-4
Tabel IV-2	: Kebutuhan Non-Fungsional Perangkat Lunak Pengujian.....	IV-5
Tabel IV-3	: Definisi Aktor.....	IV-9
Tabel IV-4	: Definisi <i>Use Case</i>	IV-9
Tabel IV-5	: Skenario <i>Use Case</i> Menambahkan Dokumen.....	IV-11
Tabel IV-6	: Skenario <i>Use Case</i> Pra-Pengolahan Dokumen.....	IV-12
Tabel IV-7	: Skenario <i>Use Case</i> Menjalankan Proses <i>POS-Tagging</i> Serial...IV-13	
Tabel IV-8	: Skenario <i>Use Case</i> Menjalankan Proses <i>Stemming</i> Serial.....	IV-14
Tabel IV-9	: Skenario <i>Use Case Split</i> Dokumen.....	IV-15
Tabel IV-10	: Skenario <i>Use Case</i> Menjalankan Proses <i>POS-Tagging</i> Paralel.....	IV-16
Tabel IV-11	: Skenario <i>Use Case</i> Menjalankan Proses <i>Stemming</i> Paralel.....	IV-17
Tabel IV-12	: Definisi Implementasi Kelas.....	IV-32
Tabel IV-13	: Perancangan Pengujian Unit.....	IV-35
Tabel IV-14	: Pengujian Sistem Pemrosesan Paralel untuk <i>Stemming</i> dan <i>POS-Tagging</i> pada Teks Bahasa Indonesia.....	IV-36

Tabel V-1 : Hasil Pengujian Tahap I-1 <i>POS-Tagging</i> dengan pemrosesan serial...	V-3
Tabel V-2 : Hasil Pengujian Tahap I-2 <i>POS-Tagging</i> dengan pemrosesan paralel alokasi 2 prosesor.....	V-4
Tabel V-3 : Reduksi Waktu Pemrosesan paralel <i>POS-Tagging</i> dengan 2 Prosesor	V-4
Tabel V-4 : Hasil Pengujian Tahap I-3 <i>POS-Tagging</i> dengan pemrosesan Paralel Alokasi 4 Prosesor.....	V-6
Tabel V-5 : Reduksi Waktu Pemrosesan paralel <i>POS-Tagging</i> dengan 4 Prosesor	V-6
Tabel V-6 : Hasil Pengujian Tahap I-4 <i>POS-Tagging</i> dengan pemrosesan Paralel Alokasi 6 Prosesor.....	V-8
Tabel V-7 : Reduksi Waktu Pemrosesan paralel <i>POS-Tagging</i> dengan 6 Prosesor	V-8
Tabel V-8 : Hasil Pengujian Tahap II-1 <i>Stemming</i> dengan pemrosesan serial.....	V-10
Tabel V-9 : Hasil Pengujian Tahap II-2 <i>Stemming</i> dengan Pemrosesan Paralel Alokasi 2 Prosesor.....	V-11
Tabel V-10 : Reduksi Waktu Pemrosesan paralel <i>Stemming</i> dengan 2 Prosesor.....	V-11
Tabel V-11 : Hasil Pengujian Tahap II-3 <i>Stemming</i> dengan Pemrosesan Paralel Alokasi 4 Prosesor.....	V-13
Tabel V-12 : Reduksi Waktu Pemrosesan paralel <i>Stemming</i> dengan 4 Prosesor.....	

.....	V-13
Tabel V-13 : Hasil Pengujian Tahap II-4 <i>Stemming</i> dengan Pemrosesan Paralel Alokasi 6 Prosesor.....	V-15
Tabel V-14 : Reduksi Waktu Pemrosesan paralel <i>Stemming</i> dengan 6 Prosesor.....	V-15
Tabel V-15 : Reduksi Waktu Proses <i>POS-Tagging</i> Tertinggi Berdasarkan Alokasi Prosesor.....	V-17
Tabel V-16 : Reduksi Waktu Proses <i>Stemming</i> Tertinggi Berdasarkan Alokasi Prosesor.....	V-17

DAFTAR GAMBAR

Gambar II-1	: Perbedaan Pemrosesan Serial dan Pemrosesan Paralel.....	II-2
Gambar II-2	: Taksonomi <i>Flynn</i>	II-2
Gambar II-3	: <i>ID</i> Proses.....	II-5
Gambar II-4	: Hasil Proses <i>Stemming</i>	II-7
Gambar II-5	: Cara Kerja Algoritma <i>Porter Stemmer</i>	II-8
Gambar II-6	: Hasil proses <i>POS-Tagging</i>	II-11
Gambar II-7	: Cara Kerja Algoritma <i>CRF-Tagger</i>	II-16
Gambar II-8	: <i>Rational Unified Process (RUP)</i>	II-17
Gambar II-9	: Hukum <i>Amdahl</i>	II-19
Gambar III-1	: Tahapan Penelitian.....	III-3
Gambar III-2	: Alur Kerja Perangkat Lunak Pengujian.....	III-6
Gambar IV-1	: Diagram <i>Use Case</i> Sistem.....	IV-8
Gambar IV-2	: <i>Activity Diagram</i> Menambahkan Dokumen.....	IV-20
Gambar IV-3	: <i>Activity Diagram</i> Pra-Pengolahan Dokumen.....	IV-21
Gambar IV-4	: <i>Activity Diagram</i> Menjalankan Proses <i>Stemming</i> Serial.....	IV-21
Gambar IV-4	: <i>Activity Diagram</i> Menjalankan Proses <i>POS-Tagging</i> Serial.....	IV-21
	IV-21
Gambar IV-5	: <i>Activity Diagram Split</i> Dokumen.....	IV-22
Gambar IV-6	: <i>Activity Diagram</i> Menjalankan Proses <i>POS-Tagging</i> Paralel.....	IV-22
	IV-22
Gambar IV-7	: <i>Activity Diagram</i> Menjalankan Proses <i>Stemming</i> Paralel.....	IV-23

Gambar IV-8	: <i>Sequence Diagram</i> Menambahkan Dokumen.....	IV-24
Gambar IV-9	: <i>Sequence Diagram</i> Menjalankan <i>POS-Tagging</i> Serial.....	IV-25
Gambar IV-10	: <i>Sequence Diagram</i> Menjalankan Proses <i>Stemming</i> Serial....	IV-26
Gambar IV-11	: <i>Sequence Diagram Split</i> Dokumen.....	IV-27
Gambar IV-12	: <i>Sequence Diagram</i> Menjalankan <i>POS-Tagging</i> Paralel.....	IV-28
Gambar IV-13	: <i>Sequence Diagram</i> Menjalankan Proses <i>Stemming</i> Paralel. .	IV-29
Gambar IV-14	: Perancangan Antarmuka Perangkat Lunak.....	IV-30
Gambar IV-15	: <i>Class Diagram</i> Pemrosesan Paralel untuk <i>Stemming</i> dan <i>POS-Tagging</i> pada Teks Bahasa Indonesia.....	IV-31
Gambar IV-16	: Implementasi Antarmuka Perangkat Lunak.....	IV-33
Gambar V-1	: Grafik Perbandingan Performansi Waktu Proses Pengujian tahap I-1 dan I-2.....	V-5
Gambar V-2	: Grafik Perbandingan Performansi Waktu Proses Pengujian tahap I-1 dan I-3.....	V-7
Gambar V-3	: Grafik Perbandingan Performansi Waktu Proses Pengujian tahap I-1 dan I-4.....	V-9
Gambar V-4	: Grafik Perbandingan Performansi Waktu Proses Pengujian tahap II-1 dan II-2.....	V-12
Gambar V-5	: Grafik Perbandingan Performansi Waktu Proses Pengujian tahap II-1 dan II-3.....	V-14
Gambar V-6	: Grafik Perbandingan Performansi Waktu Proses Pengujian tahap II-1 dan II-4.....	V-16

DAFTAR LAMPIRAN

- LAMPIRAN I : KODE PROGRAM
- LAMPIRAN II : *LOG BOOK* PENELITIAN
- LAMPIRAN III : LEMBAR PENGUJIAN PENELITIAN

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan menjelaskan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah serta sistematika penulisan sebagai gambaran umum mengenai penelitian yang akan dilakukan.

1.2 Latar Belakang

Stemming dan *POS-Tagging* merupakan bagian dari tahap pra-pengolahan data mentah teks pada bidang Pemrosesan Bahasa Alami yang bertujuan untuk menghasilkan data yang lebih terstruktur sebelum di proses ke tahap berikutnya. Hasil pengolahan pada tahap *POS-Tagging* sangat penting sebagai proses awal dalam pemrosesan bahasa untuk memberikan definisi dan konteks suatu kata (Kamayani, 2019). Dua tahap ini sering dianggap sebagai langkah awal yang sangat mempengaruhi performa proses pengaplikasian Pemrosesan Bahasa Alami (Dipongkor *et al.*, 2021).

Stemming merupakan tahap pra-pengolahan data mentah teks yang akan menghasilkan kumpulan kata dasar dari sebuah struktur kalimat. Prosesnya berupa proses pemetaan dan penguraian sehingga diperoleh kata dasar dari suatu kata yang mengalami imbuhan dengan cara menghilangkan imbuhan tersebut. Proses *Stemming* akan memberikan pemetaan varian morfologis kata yang berbeda keda-

lam kata dasar (Tala, 2003). Tujuan tahap ini adalah untuk mereduksi bentuk infleksional dan terkadang bentuk kata turunan menjadi bentuk dasar yang sama (Watzlawik and Valsiner, 2012).

POS-Tagging merupakan proses pemberian label pada setiap kata dalam kalimat dengan *POS (Part-Of-Speech)* atau *tag* yang sesuai dengan kelas kata seperti kata kerja, kata keterangan, kata sifat dan lainnya. *POS* juga disebut sebagai kategori gramatikal dari sebuah kata, yang menandakan perilaku morfologis dan sintaksis dari item leksikal (Fu *et al.*, 2018). *POS-Tagging* merupakan pemrosesan lanjutan pengolahan data mentah teks setelah tahap *Stemming*.

Data hasil pengolahan pada tahap *Stemming* dan *POS-Tagging* dapat dimanfaatkan untuk pengolahan teks dengan berdasarkan tata aturan tertentu dalam sebuah struktur tata bahasa yang memiliki struktur berbeda. Salah satunya untuk pengolahan teks Bahasa Indonesia. Namun, menurut Rianto *et al.*(2021) tingkat efisiensi waktu proses masih rendah untuk ukuran data yang besar. Akan tetapi performa waktu proses tidak hanya dipengaruhi ukuran data, teknik pemrosesan teks yang dilakukan juga akan mempengaruhi performa waktu pemrosesannya (Yusliani *et al.*, 2019).

Penelitian sebelumnya telah mengembangkan metode peningkatan performa waktu proses *Stemming* dan *POS-Tagging*. Salah satunya pengembangan metode Pemrosesan Paralel untuk mereduksi waktu pengolahan pada tahap *Stemming* dengan hasil peningkatan performa waktu proses yang signifikan (Yusliani *et al.*,

2019). Namun, metode yang dikembangkan ini masih terbatas pada peningkatan performa waktu pemrosesan untuk proses *Stemming* dan data uji yang digunakan pada pengujian penelitian ini tergolong berukuran kecil.

Pemrosesan Paralel merupakan metode yang bertujuan untuk menghasilkan proses yang akan berjalan secara bersamaan sehingga program komputer melakukan banyak proses dalam satu waktu untuk mempercepat waktu komputasi. Performa komputasi paralel terbukti menghasilkan nilai yang lebih baik dibandingkan komputasi serial (Rahmatullah *et al.*, 2021).

Maka pada penelitian ini akan mengembangkan penelitian sebelumnya dengan menerapkan metode Pemrosesan Paralel untuk tahap lanjutan setelah tahap *Stemming* yaitu pada tahap *POS-Tagging*. Selain itu, peneliti akan menerapkan metode yang telah dikembangkan untuk proses pra-pengolahan teks Bahasa Indonesia dengan ukuran data yang lebih besar untuk menguji tingkat efisiensi waktu proses yang dihasilkan.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, rumusan masalah pada penelitian ini adalah bagaimana pengaruh Pemrosesan Paralel terhadap performa waktu pemrosesan untuk tahap *Stemming* dan *POS-Tagging* teks Bahasa Indonesia dengan ukuran data yang lebih besar?

1.4 Tujuan Penelitian

Tujuan Penelitian ini adalah sebagai berikut:

1. Melakukan implementasi pemrosesan secara serial *Stemming* dan *POS-Tagging* untuk pemrosesan teks Bahasa Indonesia dengan ukuran data pada rentang 7 Mb – 25 Mb untuk proses *Stemming* dan 10 Mb – 70 Mb untuk proses *POS-Tagging*.
2. Melakukan implementasi pemrosesan secara paralel *Stemming* dan *POS-Tagging* untuk pemrosesan teks Bahasa Indonesia untuk ukuran data dengan ukuran data pada rentang 7 Mb – 25 Mb untuk proses *Stemming* dan 10 Mb – 70 Mb untuk proses *POS-Tagging*.
3. Melakukan analisis performansi dari implementasi metode Pemrosesan Serial dan Pemrosesan Paralel.

1.5 Manfaat Penelitian

Manfaat bagi pengguna adalah dapat meningkatkan efisiensi kinerja waktu proses tahap pra-pengolahan teks Bahasa Indonesia untuk ukuran yang lebih besar dan sebagai referensi bagi peneliti lain untuk tahap pengembangan lebih lanjut.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Uji coba tahap *Stemming* menggunakan pustaka Sastrawi.

2. *Corpus* yang digunakan untuk proses *POS-Tagging* adalah *IDN-Tagged Corpus* dan *POS-Tagger* yang akan digunakan adalah *CRF Tagger*.
3. Proses paralelisasi menggunakan *Python Multiprocessing Module*.
4. Penelitian ini hanya akan berfokus pada peningkatan performansi waktu proses.
5. Data uji yang digunakan adalah dokumen *plain text (*.txt)* teks Bahasa Indonesia.
6. Rentang ukuran data uji untuk proses *POS-Tagging* adalah 10 MB – 70 MB dan untuk proses *Stemming* adalah 7 MB – 25 MB.

1.7 Sistematika Penulisan

Sistematika penulisan pada karya tulis ini akan merujuk pada panduan penulisan Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini akan diuraikan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan diuraikan mengenai dasar-dasar teori yang digunakan dalam penelitian, seperti metode Pemrosesan Paralel, *Stemming*, dan *POS-Tagging*.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan diuraikan tahapan yang akan dilaksanakan pada penelitian. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Pada akhir bab, akan disertakan perancangan manajemen proyek pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan diuraikan fase-fase yang akan dijalankan pada tahapan pengembangan perangkat lunak. Fase yang dijalankan berdasarkan pada metode pengembangan yang telah ditentukan merujuk pada metodologi penelitian yang telah diuraikan pada bab sebelumnya.

BAB V. HASIL DAN PEMBAHASAN

Pada bab ini akan diuraikan hasil proses pengujian penelitian menggunakan perangkat lunak yang telah dikembangkan. Selanjutnya, hasil pengujian akan di analisis dan akan diuraikan secara detail terkait hasil-hasil yang diperoleh selama proses pengujian.

BAB VI. KESIMPULAN DAN SARAN

Bab ini merupakan bab penutup berupa penarikan kesimpulan dari semua rangkaian proses penelitian yang telah dilakukan. Selain itu, pada bab ini akan diuraikan saran-saran yang diharapkan dapat berguna untuk keperluan pengembangan penelitian kedepannya.

1.8 Kesimpulan

Pada bab pendahuluan ini, telah dikemukakan gambaran dari penelitian yang akan dilakukan. Meliputi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian serta gambaran umum terkait sistematika penulisan.

DAFTAR PUSTAKA

- Agusta, L. (2009) ‘Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia’, *Konferensi Nasional Sistem dan Informatika 2009*, (KNS&I09-036), pp. 196–201.
- A. Aziz, Z., Naseradeen Abdulqader, D., Sallow, A. B., & Khalid Omer, H. (2021). Python Parallel Processing and Multiprocessing: A Rivew. *Academic Journal of Nawroz University*, 10(3), 345–354.
<https://doi.org/10.25007/ajnu.v10n3a1145>
- Dipongkor, A.K., Nashiry, Md.A., Abdullah, K.A. and Ritu, R.S. (2021). A Study on Bengali Stemming and Parts-of-Speech Tagging. *Advances in Intelligent Systems and Computing*, [online] pp.35–44.
https://link.springer.com/chapter/10.1007/978-981-33-4367-2_4
- Fu, S. *et al.* (2018) ‘Towards Indonesian Part-of-Speech Tagging: Corpus and Models’, *Proceedings of LREC 2018 Workshop on Belt and Road LRE*, 1, pp. 2–7. <http://universaldependencies.org/>
- Jurafsky, D S. (2000). *Speech and Language Processing "An Introduction to Natural Language Processing, Cmputationak Linguistics, and Specch Reconition.* Prentice-Hall, Inc. New Jersey
- Kamayani, M. (2019) ‘Perkembangan Part-of-Speech Tagger Bahasa Indonesia’, *Jurnal Linguistik Komputasional (JLK)*, 2(2), p. 34. doi:

10.26418/jlk.v2i2.20.

Made, I. and Agastya, A. (2018). PENGARUH STEMMER BAHASA INDONESIA TERHADAP PEFORMA ANALISIS SENTIMEN TERJEMAHAN ULASAN FILM. *Jurnal TEKNOKOMPAK*, [online] 12(1), pp.18–23.

Misriyah, Anis (2011). Tata bahasa

Muhammad, A. and Sadeli, I. (2013) ‘Implementasi Dan Pengujian Algoritma Parallel Kalman Filter Pada Platform Komputer Multicore’, 2(2), pp.15–18.

Nurwidyantoro, A. and Winarko, E. (2012). *Parallelization of Maximum Entropy POS Tagging for Bahasa Indonesia with MapReduce*. [online] arXiv.org. <https://arxiv.org/abs/1208.3047>

Pisceldo, F., Adriani, M. and Manurung, R. (2009) ‘Probabilistic Part of Speech Tagging for Bahasa Indonesia’, *Proceedings of the 3rd International MALINDO Workshop, colocated event ACL-IJCNLP*, (January 2009).

Rahmatullah, G. M., Zulkarnain, A. F. and Hidayat, M. R. (no date) ‘Analisis perbandingan performa pemrograman sekuensial dan paralel dengan skema uji matrix, filter dan quick sort’, pp. 19–24.

Rajani Shree, M. and Shambhavi, B.R. (2020). POS tagger model for Kannada text with CRF++ and deep learning approaches. *Journal of Discrete Mathematical Sciences and Cryptography*, 23(2), pp.485–493. doi:10.1080/09720529.2020.1728902.

- Rianto, Mutiara, A.B., Wibowo, E.P. and Santosa, P.I. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*, [online] 8(1). <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-02100413-1>
- Sukanto, & Shalahuddin. (2013). *Analisa dan Desain Sistem Informasi*. Yogyakarta: Andi Offset.
- Tala, F. Z. (2003) 'A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia', *M.Sc. Thesis, Appendix D*, pp. 39–46.
- Watzlawik, M. and Valsiner, J. (2012) 'The Making of Magic: Cultural Constructions of the Mundane Supernatural', *The Oxford Handbook of Culture and Psychology*, 2(6), pp. 1930–1938. doi: 10.1093/oxfordhb/9780195396430.013.0038.
- Wilkinson, B. (2014) 'Permainan Catur Di Cluster Beowulf', pp. 1–6.
- Wiryananda, N. 2002. *Perancangan Perangkat Lunak Penerjemah teks Bahasa Inggris ke Teks Bahasa Indonesia*. Bandung: Tesis.
- Wu, W., Zhang, B. and Ostendorf, M. (2010) 'Automatic generation of personalized annotation tags for Twitter users', *NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*, (June), pp. 689–692.
- Yusliani, N., Primartha, R. and Diana, M. (2019) 'Multiprocessing Stemming: A

Case Study of Indonesian Stemming', *International Journal of Computer Applications*, 182(40), pp. 15–19. doi: 10.5120/ijca2019918476.