

PENGARUH POS *TAGGING* DENGAN METODE HMM PADA
NER MENGGUNAKAN *RULE BASED* UNTUK ARTIKEL
BERITA BAHASA INDONESIA

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



Oleh :

ADE LISMITA ZULKARNAIN
09021181520016

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN TUGAS AKHIR

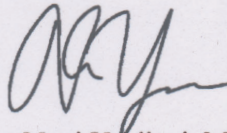
**PENGARUH POS *TAGGING* DENGAN METODE HMM PADA NER
MENGUNAKAN *RULE BASED* UNTUK ARTIKEL
BERITA BAHASA INDONESIA**

Oleh:

**ADE LISMITA ZULKARNAIN
NIM : 09021181520016**

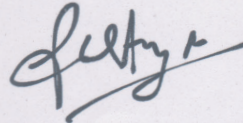
Palembang, Juli 2020

Pembimbing I,



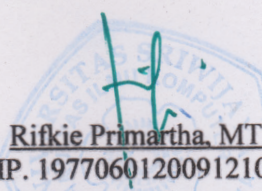
Novi Yusliani, M.T
NIP. 198211082012122001

Pembimbing II,



Desty Rodiah, M.T
NIP. 1671016112890005

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, MT
NIP. 197706012009121004



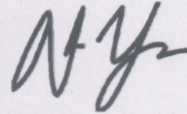
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Senin tanggal 27 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Ade Lismita Zulkarnain
NIM : 09021181520016
Judul : Pengaruh POS *Tagging* dengan Metode HMM pada NER
Menggunakan *Rule Based* untuk Artikel Berita Bahasa Indonesia

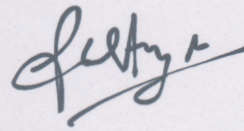
1. Pembimbing I

Novi Yusliani, M.T
NIP. 198211082012122001



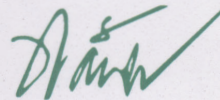
2. Pembimbing II

Desty Rodiah, M.T
NIP. 1671016112890005



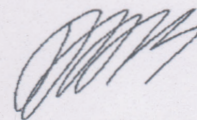
3. Penguji I

Samsuryadi, M.Kom., Ph.D.
NIP. 197102041997021003



4. Penguji II

Danny Matthew Saputra, S.T., M.Sc.
NIP. 198505102015041002



Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, MT
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Ade Lismita Zulkarnain
NIM : 09021181520016
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh POS *Tagging* dengan Metode HMM
pada NER Menggunakan *Rule Based* untuk Artikel
Berita Bahasa Indonesia

Hasil Pengecekan Software *iThenticate/Turnitin* : 7%

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.



Palembang, Juli 2020



Ade Lismita Zulkarnain
NIM. 09021181520016

MOTTO DAN PERSEMBAHAN

MOTTO:

Sesungguhnya perkataan yang paling benar adalah Kitabullah, penyelamat dari yang paling terpercaya adalah kalimat takwa, agama yang paling baik adalah agama Ibrahim, sunnah yang paling baik adalah sunnah Muhammad SAW, pembicaraan yang paling mulia adalah ingat Allah SWT, cerita yang paling baik adalah Al-Qur'an, dan surga yang paling dekat adalah orang tua.

∩

“Hidup itu sulit. Kita harus menghadapinya dengan berani“

If you believe in your dreams, I will prove to you, that you can achieve your dreams just by working hard.

온 마음으로 당신의 꿈을 이루십시오

Kupersembahkan karya tulis ini kepada :

- *Orang tuaku tercinta dan adikku tersayang*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

EFFECT OF POS TAGGING WITH HMM METHOD ON RULE BASED NER FOR INDONESIA LANGUAGE NEWS ARTICLE

By:

Ade Lismita Zulkarnain
09021181520016

ABSTRACT

Nowadays the presentation of information can be easily found in various online media, one of them is article. The Article contains various information which is arranged in a very long writing and it takes a long time to find the information. One of the fields that can process hidden information in news articles is information extraction. The main task of information extraction is named entity recognition (NER). NER aims to identify and classify the words from the documents into some categories of named-entity that have been set. In this study, NER was done in Indonesian news articles by using Rule Based to classify words into three classes of named-entity which are person, location, and organization. The rule regulations created based on contextual feature, morphology feature, and word class feature using post tagging with hidden markove model (HMM) method. This study will look at the performance result of NER system and the effect of POS tagging with HMM method againts word class feature of NER. The testing was done in 50 indonesian text news articles and the F-measure value was 62,55%.

Keywords: Named Entity Recognition, Rule Based, POS Tagging with Hidden Markove Model

PENGARUH POS *TAGGING* DENGAN METODE HMM PADA NER
MENGUNAKAN *RULE BASED* UNTUK ARTIKEL

BERITA BAHASA INDONESIA

Oleh:

Ade Lismita Zulkarnain
09021181520016

ABSTRAK

Saat ini penyajian informasi dapat dengan mudah ditemukan dalam berbagai media *online* yang salah satunya berupa artikel. Artikel memuat beragam informasi yang dikemas dalam penulisan yang sangat panjang dan membutuhkan waktu yang cukup lama untuk menemukan informasi tersebut. Salah satu bidang yang dapat memproses informasi tersembunyi pada artikel berita adalah ekstraksi informasi. Tugas utama dari ekstraksi informasi adalah *named entity recognition* (NER). NER bertujuan mengidentifikasi dan mengklasifikasikan kata dari dokumen ke dalam beberapa kategori *named-entity* yang telah ditetapkan. Dalam penelitian ini, NER dilakukan pada artikel berita bahasa Indonesia menggunakan *Rule Based* untuk mengklasifikasikan kata ke dalam tiga kelas *named-entity* yaitu *person*, *location*, dan *organization*. Aturan *rule* dibuat berdasarkan fitur kontekstual, fitur morfologi dan fitur kelas kata menggunakan POS *tagging* dengan metode *hidden markove model* (HMM). Penelitian ini akan melihat hasil kinerja sistem NER dan pengaruh POS *tagging* dengan metode HMM terhadap fitur kelas kata NER. Pengujian dilakukan pada 50 teks artikel berita bahasa Indonesia dan didapatkan nilai *F-measure* sebesar 62,55%.

Kata kunci: *Named Entity Recognition*, *Rule Based*, *POS Tagging* dengan *Hidden Markove Model*

KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku tercinta, Ayah Ahmad Zulkarnain dan Ibu Lismawati, Adikku tersayang Nabilah Zahirah dan seluruh keluarga besarku yang selalu mendoakan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika, dan Ibu Hardini Novianti, M.T selaku Sekretaris Jurusan Teknik Informatika.
3. Ibu Novi Yusliani, M.T. selaku dosen pembimbing I dan Ibu Desty Rodiah, M.T selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi serta kasih sayang kepada penulis selama proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Rifkie Primartha, M.T selaku dosen pembimbing akademik, yang telah membimbing dan selalu memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak Ir. M. Ihsan Jambak, M.Sc selaku dosen penguji I Seminar Proposal, Bapak Samsuryadi, M.Kom., Ph.D selaku dosen penguji I Komprehensif, dan Bapak Danny Matthew Saputra, S.T., M.Sc selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Aisyah, Desti, Devi Permata H, Desi, Ima, Arfah, Zakia, Tara-Tira serta adik sepupuku tersayang Riska Wati S, dan seluruh teman jurusan Teknik Informatika yang telah berbagi keluh kesah, motivasi, semangat, dan canda tawa selama masa perkuliahan.
9. Sri Astuti, SE ~ (이모)
10. Thom-tom (Muhammad Albar) ~ प्यार
11. Sahabat terbaikku Dinot (Dini S Virginanda, S.Pd), Oni Depay (Devi P Lestari, S.KG), Oni Iis (Iis Ariska, S.Psi) yang selalu meluangkan waktu untuk berbagi keluh kesah, berbagi nasihat, motivasi serta doa dan kasih sayang yang membuat penulis tidak akan pernah lupa.
12. Sahabat *Mangatoon*, *K-Group* dan *India Group* yang selalu menemani mengisi waktu luang, berbagi pengalaman, cerita dan hal menarik yang selalu menginspirasi penulis.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

dralaya, Juli 2020

Ade Lismita Zulkarnain

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN LEMBAR PENGESAHAN	ii
HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-7
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan	II-1
2.2 <i>Named Entity Recognition</i>	II-1
2.3 Praproses Teks	II-4
2.4 <i>Hidden Markov Model POS Tagging</i>	II-5
2.4.1 Algoritma Viterbi	II-9
2.5 <i>Rule Based</i>	II-11

2.6	Evaluasi	II-13
2.7	Artikel	II-19
2.8	<i>Rational Unified Process</i> (RUP)	II-20
2.9	Penelitian Lain yang Relevan	II-21
2.10	Kesimpulan	II-24

BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan	III-1
3.2	Metode Pengumpulan Data	III-1
3.3	Tahapan Penelitian	III-1
3.3.1	Menetapkan Kerangka Kerja	III-2
3.3.1.1	Praproses	III-3
3.3.1.2	<i>Rule Assignment</i>	III-13
3.3.1.3	Evaluasi Kinerja Sistem Hasil NER	III-15
3.3.2	Menetapkan Kriteria Pengujian	III-15
3.3.3	Menetapkan Format Data Pengujian	III-16
3.3.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-16
3.3.5	Melakukan Pengujian Penelitian	III-17
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-17
3.4	Metode Pengembangan Perangkat Lunak	III-18
3.4.1	Fase Insepsi	III-18
3.4.2	Fase Elaborasi	III-19
3.4.3	Fase Konstruksi	III-19
3.4.4	Fase Transisi	III-20
3.5	Manajemen Proyek Penelitian	III-20

BAB IV METODOLOGI PENELITIAN

4.1.	Pendahuluan	IV-1
4.2.	<i>Rational Unified Process</i> (RUP)	IV-1

4.2.1	Fase Insepsi	IV-1
4.2.1.1	Pemodelan Bisnis	IV-1
4.2.1.2	Kebutuhan Sistem	IV-2
4.2.1.3	Analisis dan Desain	IV-4
4.2.1.3.1	Analisis Kebutuhan Perangkat Lunak.....	IV-4
4.2.1.3.2	Analisis Data	IV-5
4.2.1.3.3	Analisis Praproses Data	IV-9
4.2.1.3.4	Analisis POS <i>Tagging</i>	IV-11
4.2.1.3.5	Analisis <i>Rule Based</i>	IV-23
4.2.1.3.6	Desain Perangkat Lunak	IV-27
4.2.2	Fase Elaborasi	IV-33
4.2.2.1	Pemodelan Bisnis	IV-33
4.2.2.1.1	Perancangan Data.....	IV-33
4.2.2.1.2	Perancangan Antarmuka	IV-35
4.2.2.2	Kebutuhan Sistem	IV-38
4.2.2.3	Diagram	IV-39
4.2.2.3.1	Diagram Aktivitas	IV-39
4.2.2.3.2	Diagram Alur	IV-41
4.2.3	Fase Konstruksi	IV-45
4.2.3.1	Kebutuhan Sistem	IV-45
4.2.3.2	Diagram Kelas	IV-45
4.2.3.3	Kelas Analisis	IV-47
4.2.3.4	Implementasi	IV-48
4.2.3.3.1	Implementasi Kelas	IV-48
4.2.3.3.2	Implementasi Antarmuka	IV-53
4.2.4	Fase Transisi	IV-54
4.2.4.1	Pemodelan Bisnis	IV-54
4.2.4.2	Kebutuhan Sistem	IV-55
4.2.4.3	Rencana Pengujian	IV-55
4.2.4.3.1	Rencana Pengujian <i>Use Case</i> Melakukan Memilih <i>File</i>	IV-56

4.2.4.3.2	Rencana Pengujian <i>Use Case</i> Melakukan Proses NER	IV-56
4.2.4.4	Implementasi	IV-57
4.2.4.4.1	Pengujian <i>Use Case</i> Melakukan Memilih <i>File</i>	IV-58
4.2.4.4.2	Pengujian <i>Use Case</i> Melakukan Proses NER	IV-59
4.2.4.4.3	Pengujian <i>Use Case</i> Melakukan POS <i>Tagging</i> dengan HMM	IV-60
4.3.	Kesimpulan.....	IV-61
 BAB V HASIL DAN ANALISIS PENELITIAN		
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Hasil Pengujian	V-3
5.3	Analisis Hasil Penelitian	V-7
5.4	Kesimpulan	V-9
 BAB VI KESIMPULAN DAN SARAN		
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran	VI-2
DAFTAR PUSTAKA		xx
LAMPIRAN		xxii

DAFTAR TABEL

	Halaman
II-1. Tabel <i>Result of Tokenization and Feature Assignment Processes</i>	II-12
II-2. Tabel Kode Hasil Perbandingan Objek <i>Named-Entity</i>	II-14
III-1. Tabel Atribut Dasar Token	III-4
III-2. Tabel Jenis Token	III-5
III-3. Tabel Fitur Kontekstual	III-7
III-4. Tabel Fitur Morfologi	III-10
III-5. Tabel Daftar <i>Tagset</i> Bahasa Indonesia	III-12
III-6. Tabel Atribut Token yang Dipengaruhi pada Tahap <i>Rule Assignment</i>	III-14
III-7. Tabel Rancangan Tabel Hasil Uji NER Berdasarkan <i>Score Report</i> ...	III-16
III-8. Tabel Rancangan Tabel Hasil Analisis Klasifikasi	III-18
III-9. Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure (WBS)</i>	III-21
IV.1 Tabel Kebutuhan Fungsional	IV-3
IV.2 Tabel Kebutuhan Non-Fungsional	IV-4
IV.3 Tabel Label <i>Named-Entity</i>	IV-7
IV.4 Tabel Nilai Probabilitas Awal dari Contoh Data <i>Training POS Tagging</i>	IV-12
IV.5 Tabel Hasil Perhitungan Probabilitas Transisi dari Contoh Data <i>Training POS Tagging</i>	V-13
IV.6 Tabel Hasil Perhitungan Probabilitas Emisi dari Contoh Data <i>Training POS Tagging</i>	IV-14
IV.7 Tabel Definisi Aktor	IV-28
IV.8 Tabel Definisi <i>Use Case</i>	IV-28
IV.9 Tabel Skenario <i>Use Case</i> Memilih <i>File</i>	IV-30
IV.10 Tabel Skenario <i>Use Case</i> Melakukan Proses NER	IV-31
IV.11 Tabel Skenario <i>Use Case</i> Melakukan <i>POS Tagging</i> dengan HMM ...	IV-32
IV.12 Tabel Implementasi Kelas	IV-49
IV.13 Tabel Rencana Pengujian <i>Use Case</i> Memilih <i>File</i>	IV-56
IV.14 Tabel Rencana Pengujian <i>Use Case</i> Melakukan Proses NER	IV-56

IV.15	Tabel Pengujian <i>Use Case</i> Memilih <i>File</i>	IV-58
IV.16	Tabel Pengujian <i>Use Case</i> Melakukan Proses NER	IV-59
IV.17	Tabel Pengujian <i>Use Case</i> Melakukan POS <i>Tagging</i> dengan HMM	IV-60
V.1	Tabel <i>Score Report</i> Hasil Dokumen Uji NER Menggunakan Fitur POS <i>Tagging</i> dengan Metode HMM	V-5
V.2	Tabel Hasil Evaluasi Pengujian Sistem NER Menggunakan Fitur POS <i>Tagging</i> dengan Metode HMM	V-5
V.3	Tabel <i>Score Report</i> Hasil Dokumen Uji NER Menggunakan POS <i>Tagging</i> tanpa HMM	V-6
V.4	Tabel Hasil Evaluasi Pengujian Sistem NER Menggunakan Fitur POS <i>Tagging</i> tanpa HMM	V-6

DAFTAR GAMBAR

	Halaman
II-1. Gambar Contoh Penggunaan NER pada Suatu Kalimat	II-3
II-2. Gambar Contoh POS <i>Tagging</i>	II-5
II-3. Gambar Bagan Proses Evaluasi (MUC, 2001)	II-13
II-4. Gambar Rumusan Parameter Kinerja Sistem NER	II-16
II-5. Gambar Contoh Dokumen <i>Key</i>	II-16
II-6. Gambar Contoh Dokumen <i>Response</i>	II-16
II-7. Gambar Arsitektur Model Iteratif RUP (Kruchten, 2004)	II-20
III-1. Gambar Diagram Tahapan Proses Perangkat Lunak Sistem NER Menggunakan Fitur POS <i>Tagging</i> dengan HMM	III-2
III-2. Gambar Diagram Tahapan Proses Perangkat Lunak Sistem NER Menggunakan Fitur POS <i>Tagging</i> tanpa HMM.....	III-3
III-3. Gambar Tahapan Pengujian Penelitian	III-17
III-4. Gambar Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-25
III-5. Gambar Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian.....	III-26
III-6. Gambar Penjadwalan untuk Tahap Menentukan Kriteria Pengujian	III-26
III-7. Gambar Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Insepsi	III-27
III-8. Gambar Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Elaboras	III-27
III-9. Gambar Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Konstruksi	III-28
III-10. Gambar Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Transisi	III-28
III-11. Gambar Penjadwalan untuk Tahap Melakukan Pengujian Penelitian	III-29
III-12. Gambar Penjadwalan untuk Tahap Analisa Hasil Pengujian Penelitian dan Membuat Kesimpulan	III-29

IV.1	Gambar Dokumen <i>Source/Response</i> atau Dokumen Uji	IV-6
IV.2	Gambar Dokumen <i>Key</i>	IV-8
IV.3	Gambar Dokumen <i>Response</i>	IV-8
IV.4	Gambar Hasil <i>Tokenization</i>	IV-10
IV.5	Gambar Contoh Data <i>Training POS Tagging</i>	IV-11
IV.6	Gambar Contoh Perhitungan Probabilitas Transisi	IV-12
IV.7	Gambar Contoh Perhitungan Probabilitas Emisi	IV-14
IV.8	Gambar Langkah Inisialisasi pada Algoritma Viterbi	IV-18
IV.9	Gambar Langkah Iteratif pada Algoritma Viterbi	IV-20
IV.10	Gambar Langkah Terminasi pada Algoritma Viterbi	IV-21
IV.11	Gambar Langkah Lacak Balik pada Algoritma Viterbi	IV-22
IV.12	Gambar Diagram <i>Use Case</i>	IV-27
IV.13	Gambar Rancangan Antarmuka Perangkat Lunak	IV-35
IV.14	Gambar Rancangan <i>File</i>	IV-35
IV.15	Gambar Kotak/ <i>Filed</i> yang Menampilkan Isi <i>File</i> yang Digunakan untuk Penelitian	IV-36
IV.16	Gambar Rancangan Fitur <i>POS Tagging</i> dengan Metode HMM	IV-36
IV.17	Gambar Rancangan Melakukan Proses NER	IV-37
IV.18	Gambar Rancangan Nilai <i>Score Report</i> Sistem NER	IV-37
IV.19	Gambar Rancangan Fitur Evaluasi Hasil Sistem NER	IV-38
IV.20	Gambar Kotak/ <i>Field</i> yang Menampilkan Isi Dokumen <i>Source</i> dengan <i>Named-Entity Tag</i> yang Berhasil Dikenali Oleh Sistem	IV-38
IV.21	Gambar Diagram Aktivitas Memilih <i>File</i>	IV-40
IV.22	Gambar Diagram Aktivitas Melakukan Proses NER	IV-40
IV.23	Gambar Diagram Aktivitas Melakukan <i>POS Tagging</i> dengan HMM	IV-41
IV.24	Gambar Diagram Alur Memilih <i>File</i>	IV-42
IV.25	Gambar Diagram Alur Melakukan Proses NER	IV-43
IV.26	Gambar Diagram Alur Melakukan <i>POS Tagging</i> dengan HMM	IV-44
IV.27	Gambar Diagram Kelas Perangkat Lunak	IV-46
IV.28	Gambar Kelas Analisis Memilih <i>File</i>	IV-47
IV.29	Gambar Kelas Analisis Melakukan Proses	IV-47

IV.30	Gambar Kelas Analisis Melakukan POS <i>Tagging</i> dengan HMM	IV-48
IV.31	Gambar Antarmuka Perangkat Lunak	IV-54
V.1	Gambar Grafik Perbandingan Hasil Pengujian NER	V-7
V.2	Gambar Perbandingan Hasil Evaluasi Sistem NER	V-8

DAFTAR LAMPIRAN

	Halaman
1. Contoh Dokumen Masukan	L-1
2. Contoh Dokumen Keluaran	L-2
3. Contoh Dokumen Keluaran 2	L-4
4. Contoh Dokumen Kunci (<i>Key</i>)	L-6
5. <i>Source Code</i> Program	L-8
6. Aturan Pengenalan <i>Named-Entity</i> Berdasarkan Fitur Kontekstual, Morfologi, dan Kelas Kata	L-54

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah. Pada bab ini juga akan memberikan penjelasan umum mengenai keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan singkat mengenai apa yang menjadi latar belakang dari penelitian ini.

1.2 Latar Belakang Masalah

Berita adalah sebuah informasi yang menjadi topik hangat dikalangan masyarakat mengenai suatu hal yang sedang terjadi. Dalam sebuah artikel berita banyak terdapat informasi penting seperti nama orang, nama tempat, tanggal, tahun, bulan, nama organisasi dan lain sebagainya. Untuk mendapatkan informasi-informasi tersebut seseorang harus membaca keseluruhan dari isi berita. Hal ini akan membutuhkan waktu yang sangat lama jika berita disajikan dalam bentuk teks yang sangat panjang. Oleh karena itu dibutuhkan proses ekstraksi informasi untuk mengambil bagian terpenting dari sebuah berita.

Ekstraksi Informasi atau *Information Extraction* (IE) merupakan salah satu bidang dalam pemrosesan bahasa alami yang bertujuan untuk menemukan informasi terstruktur dari dokumen yang tidak terstruktur (Setiyoadi *et al.*, 2017). Salah satu proses terpenting dalam ekstraksi informasi adalah mengenali entitas kata. Untuk dapat mengenali entitas kata pada kalimat dibuatlah pengenalan

entitas bernama atau *Named Entity Recognition* (NER) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan kata dari dokumen ke dalam beberapa kategori *named-entity* yang telah ditetapkan seperti nama orang, nama tempat, nama organisasi dan sebagainya (Wulandari *et al.*, 2018).

NER pertama kali digunakan ke dalam *Message Understanding Conference-6* (MUC-6) pada tahun 1995. Penerapan NER juga banyak dilakukan pada aplikasi pemrosesan bahasa alami seperti sistem tanya jawab otomatis, pengindeksan dokumen, mesin penerjemah, pencarian informasi, dan peringkasan teks (Taufik *et al.*, 2016). Beberapa pendekatan yang digunakan di dalam NER yang akan dibahas salah satunya adalah *Rule Based*. Pendekatan *rule based* pada NER dilakukan dengan menggunakan aturan atau *rule* yang dibuat berdasarkan pola yang ditemukan pada dokumen.

Penelitian mengenai NER telah banyak dilakukan. Salah satunya, Budi *et al.* (2005) melakukan penelitian *Named Entity Recognition* untuk Bahasa Indonesia menggunakan pendekatan *Rule Based*. Penggunaan *rule* pada penelitian ini dapat secara efektif mengenali suatu entitas karena didasarkan atas aturan-aturan yang menggabungkan kontekstual, morfologi, dan fitur *part of speech*. Hasil dari penelitian ini menghasilkan nilai *recall* sebesar 63.43% dan *precision* sebesar 71.84%. Wulandari *et al.* (2018) melakukan penelitian *Named Entity Recognition* menggunakan *Rule Based* dan *Naive Bayes Classifier* pada dokumen Biologi. Penelitian ini memanfaatkan *POS tagging* untuk mencari kelas kata dan menggunakan *rule* untuk menentukan *named-entity* dari setiap kata. Hasil dari penelitian ini menghasilkan nilai *recall* sebesar 0,66 dan *f-measure* sebesar 0,45.

Part of Speech (POS) Tagging merupakan proses pelabelan secara otomatis pada suatu kata dalam kalimat. Dikutip dari Aryoyudanta *et al.* (2016) bahwa proses *POS tagging* dapat memberikan informasi kelas kata yang sangat membantu dalam proses NER karena proses ini mengarahkan sistem NER untuk menemukan suatu entitas penting. Selanjutnya, proses untuk menemukan entitas kata dapat menggunakan *rule*. *Rule* bertujuan mengelompokkan kelas kata ke dalam *named-entity* yang dibuat berdasarkan aturan. Adapun aturan yang digunakan seperti sintaksis penulisan, fitur ortografi dan kelas kata atau *part of speech*. Oleh karena itu *POS tagging* memiliki pengaruh terhadap *rule based* dikarenakan *rule based* memiliki aturan berupa pelabelan yang salah satunya dapat dibuat menggunakan *part of speech*. *POS Tagging* dapat dilakukan secara manual dengan menggunakan bantuan ahli bahasa dan secara otomatis dapat menggunakan metode probabilistik, salah satunya dengan teknik *Hidden Markov Model (HMM)*. Hal ini dikarenakan proses *POS tagging* bisa dipandang sebagai proses klasifikasi suatu urutan *tag* untuk tiap kata dalam suatu kalimat. Penelitian *POS tagging* untuk teks berbahasa Indonesia menggunakan HMM telah dilakukan Wicaksono & Purwarianti (2010). Penelitian ini menghasilkan nilai akurasi yang cukup tinggi yaitu 96,5%.

Berdasarkan uraian tersebut, penelitian ini akan melakukan NER menggunakan *Rule Based* yang telah dilakukan pada penelitian Budi *et al.* (2005) untuk artikel berita Bahasa Indonesia, dimana kelas kata pada penelitian ini akan menggunakan *POS Tagging* dengan metode HMM. Penerapan *POS tagging*

menggunakan HMM pada NER ini diharapkan dapat membantu proses *rule* agar lebih mudah menemukan *named-entity* pada teks berita.

1.3 Rumusan Masalah

Masalah yang diangkat pada penelitian ini adalah bagaimana pengaruh POS *Tagging* dengan metode HMM pada NER menggunakan *Rule Based* untuk artikel berita Bahasa Indonesia?

Untuk menyelesaikan permasalahan tersebut maka diuraikan pertanyaan penelitian sebagai berikut:

1. Bagaimana penggunaan metode HMM untuk POS *Tagging* dan *Rule Based* pada proses NER ?
2. Apakah penggunaan POS *Tagging* dengan metode HMM memiliki pengaruh terhadap hasil akurasi pada poses NER ?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui penggunaan metode HMM untuk POS *Tagging* dan *Rule Based* pada proses NER.
2. Mengetahui pengaruh HMM terhadap POS *Tagging* pada proses NER.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Dapat memahami teknik HMM sebagai metode POS *Tagging*.
2. Dapat mengetahui pengaruh POS *Tagging* dengan metode HMM pada NER melalui nilai akurasi sehingga dapat dijadikan rujukan dibidang *Named Entity Recognition*.

1.6 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa artikel berita berbahasa Indonesia yang telah diubah ke dalam format **.txt*
2. Pengenalan entitas dilakukan per kata dalam kalimat.
3. *Named-entity* yang akan dikenali berupa nama orang, nama lokasi dan nama organisasi.
4. Penelitian ini tidak memperhatikan ambiguitas kata.

1.7 Sistematika Penulisan

Sistematika penulisan ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Bab I menguraikan latar belakang dan rumusan masalah, tujuan dan manfaat penulisan, batasan masalah, dan sistematika penulisan penelitian.

BAB II. KAJIAN LITERATUR

Bab II berisi landasan teori yang digunakan pada penelitian ini seperti, *Named Entity Recognition*, *POS Tagging* dengan HMM, dan *Rule Based*. Selain itu bab ini juga membahas mengenai penelitian yang terkait dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab III berisi tentang pembahasan mengenai tahapan serta pengembangan perangkat lunak yang akan dilaksanakan pada penelitian ini. Arsitektur mengenai tahapan penelitian dibangun dengan suatu sistem kerangka kerja yang akan dideskripsikan secara rinci, dan pada akhir bab akan dijabarkan mengenai manajemen proyek penelitian berupa penjadwalan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab IV membahas perancangan perangkat lunak yang akan dibangun pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab V menampilkan hasil pengujian perangkat lunak berdasarkan langkah-langkah yang telah direncanakan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab VI berisi tentang kesimpulan dari semua uraian pada semua bab dan juga berisi saran-saran yang diharapkan dapat membantu dalam pengembangan penelitian selanjutnya.

1.8 Kesimpulan

Bab ini telah menguraikan pembahasan mengenai latar belakang masalah penelitian dalam proses *Named Entity Recognition* (NER). Dimana proses terpenting dari NER adalah menemukan *named-entity* kata dalam kalimat. Penelitian ini akan berfokus untuk melihat pengaruh POS *Tagging* dengan metode HMM terhadap proses NER menggunakan *Rule Based* untuk artikel berita Bahasa Indonesia. Untuk melihat adanya pengaruh HMM pada POS *tagging* terhadap NER, maka hasil pengujian POS *tagging* menggunakan HMM akan dibandingkan dengan POS *tagging* tanpa menggunakan HMM.

DAFTAR PUSTAKA

- Aryoyudanta, B., Adji, T. B., & Hidayah, I. (2017). Semi-supervised learning approach for Indonesian Named Entity Recognition (NER) using co-training algorithm. *Proceeding - 2016 International Seminar on Intelligent Technology and Its Application, ISITIA 2016: Recent Trends in Intelligent Computational Technologies for Sustainable Energy*, 7–12. <https://doi.org/10.1109/ISITIA.2016.7828624>
- Budi, I., Hasibuan, Z. A., & Nazief, B. (2005). *Discovery Science*, 5808(October). <https://doi.org/10.1007/978-3-642-04747-3>
- Grishman, R., Sundheim, B., Road, G., Diego, S., & York, N. (n.d.). *Message Understanding Conference-6: A Brief History Ocean Surveillance Center Evaluation Division (NRaD) Short-term subtasks Portability. New York.* <https://doi.org/10.3115/992628.992709>
- Haq, A. A. (2013). Named Entity Recognition In Crime News Articles Using Hidden Markov Model. *Universitas Sriwijaya*
- Juditha, C. (2013). News Accuracy in Online Journalism (News of Alleged Corruption The Constitutional Court in Detiknews). *Jurnal Pekomnas*. 16(3): 145–154.
- Krutchén, Philippe.2003. *The Rational Unified Process: An Introduction.* Addison-Wesley.
- Kurniawati, Y., & Adikara, P. P. (2018). Implementasi Named Entity Recognition Pada Factoid Question Answering System Untuk Cerita Rakyat Indonesia. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(9), 3142–3149. Retrieved from <http://puslit2.petra.ac.id/ejournal/index.php/inf/article/view/1647>
- MUC. 2001. The Message Understanding Conference Scoring Software User's Manual [online]. Dari: http://www.itl.nist.gov/iaui/894.02/related_projects/muc/muc_sw/muc_sw_manual.html; Internet; diakses 11 Agustus 2019.
- Nithya, S. Vijayarani, & Ilamathi, J. (2018). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16. <https://doi.org/10.1016/j.procs.2013.05.286>
- Saefulloh, F. (2017). PART OF SPEECH TAGGER UNTUK BAHASA INDONESIA DENGAN MENGGUNAKAN MODIFIKASI BRILL, 9(1),

37–42.

- Setiyoaji, A., Muflikhah, L., & Fauzi, M. A. (2017). Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTTIK) Universitas Brawijaya*, 1(12), 1858–1864.
- Shalan, K. (2010). Rule-based Approach in Arabic Natural Language Processing. *International Journal on Information and Communication Technologies*, 3(3), 11–19.
- Taufik, N., Wicaksono, A. F., & Adriani, M. (2017). Named entity recognition on Indonesian microblog messages. *Proceedings of the 2016 International Conference on Asian Language Processing, IALP 2016*, 358–361. <https://doi.org/10.1109/IALP.2016.7876005>
- Wicaksono, A. F., & Ayu, P. (2010). HMM Based Part-of-Speech Tagger for Bahasa Indonesia. *Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop*, (June), 1–7. <https://doi.org/10.1186/1471-2377-12-75>
- Wulandari, D. W., Adikara, P. P., & Adinugroho, S. (2018). Named Entity Recognition (NER) Pada Dokumen Biologi Menggunakan Rule Based dan Naïve Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTTIK) Universitas Brawijaya*, 2(11), 4555–4563.