

NAMED-ENTITY RECOGNITION PADA TEKS BERBAHASA  
INDONESIA MENGGUNAKAN METODE HIDDEN MARKOV  
MODEL DAN PART-OF-SPEECH TAGGING

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



Oleh :

M. Ridho Putra Sufa  
NIM: 09021281621053

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

2020

## LEMBAR PENGESAHAN TUGAS AKHIR

### NAMED-ENTITY RECOGNITION PADA TEKS BERBAHASA INDONESIA MENGGUNAKAN METODE HIDDEN MARKOV MODEL DAN PART-OF-SPEECH TAGGING

Oleh:

M. Ridho Putra Sufa  
NIM : 09021281621053

Pembimbing I



Novi Yusufiani, S.Kom., M.T.  
NIP. 198211082012122001


Palembang, Agustus 2020  
Pembimbing II



Muhammad Ali Buchari, S.Kom., M.T.  
NIP. 198803302019031007

Mengetahui,

Ketua Jurusan Teknik Informatika




Rifkie Primartha, M.T.  
NIP. 197706012009121004

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 04 Agustus 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : M. Ridho Putra Sufa  
NIM : 09021281621053  
Judul : *Named Entity Recognition* pada Teks Berbahasa Indonesia menggunakan Metode *Hidden Markov Model* dan *Part-of-Speech Tagging*

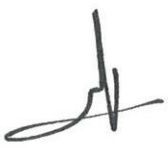
1. Pembimbing I

  
.....  
Novi Yusliani, S.Kom., M.T.  
NIP. 198211082012122001

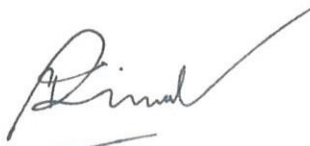
2. Pembimbing II

  
.....  
Muhammad Ali Buchari, S.Kom., M.T.  
NIP. 198803302019031007


3. Penguji I

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

4. Penguji II

  
.....  
Mastura Diana Marieska, S.T., M.T.  
NIP. 198603212018032001

Mengetahui,  
Ketua Jurusan Teknik Informatika

  
.....  
Rifkie Primartha, M.T.  
NIP. 197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : M. Ridho Putra Sufa  
NIM : 09021281621053  
Program Studi : Teknik Informatika  
Judul Skripsi : *Named Entity Recognition* pada Teks Berbahasa Indonesia menggunakan Metode *Hidden Markov Model* dan *Part-of-Speech Tagging*  
Hasil Pengecekan Software *iThenticate/Turnitin* : 10 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek 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, Agustus 2020



M. Ridho Putra Sufa  
NIM. 09021281621053

## MOTTO DAN PERSEMBAHAN

MOTTO:

*“I Think, therefore I am”*

- René Descartes

*“Even though the future seems far away, it is actually beginning right now”*

- Mattie Stepanek

*“Radiant of New Hope”*

- Penulis

*Kupersembahkan karya tulis ini kepada :*

- *Allah SWT & Nabi Muhammad SAW*
- *Ibu & ayahku tercinta*
- *Keluarga besarku*
- *Dosen pembimbing & dosen penguji*
- *Sahabat & teman seperjuanganku*
- *Almamater*

NAMED ENTITY RECOGNITION IN INDONESIAN LANGUAGE TEXT  
USING HIDDEN MARKOV MODEL AND PART-OF-SPEECH TAGGING

By:

M. Ridho Putra Sufa  
09021281621053

ABSTRACT

Named-Entity Recognition (NER) is one of the research topics in the field of Natural Language Processing (NLP), it aims to provide Named-Entity labels that refer to person's name, location name, organization name, and unit of time in the text. In this study, NER was conducted on Indonesian language text using Hidden Markov Model (HMM) method, and assisted by the Part-of-Speech Tagging process. The POS-Tagging process is applied to help HMM method in labeling Named-Entity labels on unknown words. Training data and test data used are data that has been developed by Fachri (2014) and (Syaifudin, 2016). The test was conducted on Indonesian language text with a total of 511 sentences. HMM and POS-Tagging method were successfully used to carry out the NER process, with performance level achieved based on average recall value is 83.82%, average precision value is 89.31%, and the average value of f-measure is 86.14%.

Keywords: Named Entity Recognition (NER), Natural Language Processing (NLP), Hidden Markov Model (HMM), Part-of-Speech Tagging

Supervisor I,



Novi Yusliani, S.Kom., M.T.  
NIP. 198211082012122001

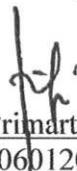
Palembang, August 2020

Supervisor II,



Muhammad Ali Buchari, S.Kom., M.T.  
NIP. 198803302019031007

Approved,  
Head of the Informatics Engineering Department



Rifkie Primartha, M.T.  
NIP. 197706012009121004



*NAMED ENTITY RECOGNITION* PADA TEKS BERBAHASA INDONESIA  
MENGUNAKAN *HIDDEN MARKOV MODEL*  
DAN *PART-OF-SPEECH TAGGING*

Oleh:

M. Ridho Putra Sufa  
09021281621053

ABSTRAK

*Named-Entity Recognition* (NER) adalah salah satu topik penelitian pada bidang Pemrosesan Bahasa Alami (PBA) yang bertujuan untuk memberikan label entitas kata yang merujuk pada nama orang, nama lokasi, nama organisasi, dan nama satuan waktu pada teks. Pada penelitian ini, NER dilakukan pada teks berbahasa Indonesia dengan menggunakan metode *Hidden Markov Model* (HMM), dan dibantu oleh proses *Part-of-Speech Tagging*. Proses *POS-Tagging* dilakukan untuk membantu metode HMM dalam melakukan pelabelan NER pada kata yang tidak diketahui. Data latih dan data uji yang digunakan yaitu data yang telah dikembangkan oleh Fachri (2014) dan (Syarifudin, 2016). Pengujian dilakukan pada teks berbahasa Indonesia dengan jumlah kalimat sebanyak 511 kalimat. Metode HMM dan *POS-Tagging* berhasil digunakan untuk melakukan proses NER, dengan tingkat performa yang dicapai yaitu nilai rata-rata *recall* sebesar 83.82%, nilai rata-rata *precision* sebesar 89.31%, dan nilai rata-rata *f-measure* sebesar 86.14%.

Kata kunci: *Named Entity Recognition* (NER), *Pemrosesan Bahasa Alami* (PBA),  
*Hidden Markov Model* (HMM), *Part-of-Speech Tagging*

Palembang, Agustus 2020

Pembimbing I,



Novi Yusliana, S.Kom., M.T.  
NIP. 198211082012122001

Pembimbing II,



Muhammad Ali Buchari, S.Kom., M.T.  
NIP. 198803302019031007

Mengetahui,  
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.  
NIP. 197706012009121004

## **KATA PENGANTAR**

Puji syukur kepada Allah 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. Ibuku tercinta, Fatmawati, yang selalu mendoakan untuk kelancaranku dalam menyelesaikan Tugas Akhir, serta yang selalu memberikan dukungan terbaik secara moril maupun materil.
2. Ayahku tercinta, Suryadi Eviontri, yang selalu mendoakan untuk kelancaranku dalam menyelesaikan Tugas Akhir, serta yang selalu memberikan dukungan terbaik secara moril maupun materil.
3. Keluarga besarku, mama Murniati, papa Jakun, om Jay, tante Leni, mbak Rani, kak Sangga, dan kak Andro yang telah memberikan dukungan yang terbaik kepadaku, dan membantuku secara moril maupun materil.
4. Pacarku tercinta, Soraya, yang selalu mendoakanku, dan selalu memberikan dukungan dan motivasi yang terbaik kepadaku disetiap harinya sehingga dapat menyelesaikan Tugas Akhir dengan baik.
5. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika, dan Ibu Alvi Syahrini Utami, M.Kom selaku Sekretaris Jurusan Teknik Informatika.
6. Ibu Novi Yusliani, S.Kom., M.T. selaku dosen pembimbing I dan Bapak Muhammad Ali Buchari, S.Kom., M.T. selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses pengerjaan Tugas Akhir, hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.



7. Bapak Muhammad Ali Buchari, S.Kom., M.T. selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
8. Bapak Abdiansah, S.Kom., M.Cs. selaku dosen penguji I, dan Ibu Mastura Diana Marieska, S.T., M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
9. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Mbak Wiwin dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
11. Adi, Melvin, Luthfi, Satria, Adit, Kak Saniyah dan Kak Achi yang telah membantu, memberikan informasi dan referensi selama proses pengerjaan Tugas Akhir.
12. Wibi, Melvin, Adi, Ganesha, serta seluruh teman jurusan Teknik Informatika yang telah berbagi keluh kesah, motivasi, semangat, dan hiburan selama masa perkuliahan.

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.

Palembang, Agustus 2020

M. Ridho Putra Sufa

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR .....	ii
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.....	xviii
BAB I PENDAHULUAN.....	I-1
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang Masalah .....	I-1
1.3    Rumusan Masalah .....	I-5
1.4    Tujuan Penelitian.....	I-6
1.5    Manfaat Penelitian.....	I-6
1.6    Batasan Masalah.....	I-6
1.7    Sistematika Penulisan.....	I-7
1.8    Kesimpulan.....	I-8
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori .....	II-1
2.2.1.    Pemrosesan Bahasa Alami.....	II-1
2.2.2. <i>Named-Entity Recognition (NER)</i> .....	II-2
2.2.3.    Pra-pemrosesan Teks .....	II-7

2.2.4.	Part-of-Speech <i>Tagging</i> .....	II-9
2.2.5.	Hidden Markov Model (HMM).....	II-13
2.2.6.	Algoritma Viterbi.....	II-14
2.2.7.	<i>Word Features</i> .....	II-16
2.2.8.	<i>Confusion Matrix</i> .....	II-19
2.2.9.	<i>Rational Unified Process (RUP)</i> .....	II-22
2.2.10.	<i>Unified Modelling Language (UML)</i> .....	II-24
2.3	Penelitian Lain yang Relevan.....	II-25
2.4	Kesimpulan.....	II-28
<b>BAB III METODOLOGI PENELITIAN.....</b>		<b>III-1</b>
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-2
3.3.1.	Kerangka Kerja.....	III-3
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-9
3.3.5.	Pengujian Penelitian .....	III-10
3.3.6.	Analisis Hasil Pengujian dan Membuat Kesimpulan .....	III-10
3.4.	Metode Pengembangan Perangkat Lunak .....	III-11
3.4.1.	Fase Insepsi.....	III-12
3.4.2.	Fase elaborasi.....	III-12
3.4.3.	Fase konstruksi .....	III-13
3.4.4.	Fase transisi .....	III-14
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK .....</b>		<b>IV-1</b>
4.1.	Pendahuluan .....	IV-1
4.2.	Fase Insepsi .....	IV-1
4.2.1.	<i>Business Modelling</i> .....	IV-1

4.2.2.	<i>Requirement</i> .....	IV-4
4.2.3.	<i>Analysis &amp; Design</i> .....	IV-7
4.2.4.	<i>Implementation</i> .....	IV-37
4.2.5.	<i>Test</i> .....	IV-37
4.3.	Fase Elaborasi.....	IV-37
4.3.1.	<i>Business Modelling</i> .....	IV-38
4.3.2.	<i>Requirement</i> .....	IV-39
4.3.3.	<i>Analysis &amp; Design</i> .....	IV-40
4.3.4.	<i>Implementation</i> .....	IV-50
4.3.5.	<i>Test</i> .....	IV-50
4.4.	Fase Konstruksi .....	IV-50
4.4.1.	<i>Business Modelling</i> .....	IV-50
4.4.2.	<i>Requirement</i> .....	IV-54
4.4.3.	<i>Analysis &amp; Design</i> .....	IV-54
4.4.4.	<i>Implementation</i> .....	IV-57
4.4.5.	<i>Test</i> .....	IV-58
4.5.	Fase Transisi .....	IV-58
4.5.1.	<i>Business Modelling</i> .....	IV-58
4.5.2.	<i>Requirement</i> .....	IV-59
4.5.3.	<i>Analysis &amp; Design</i> .....	IV-59
4.5.4.	<i>Implementation</i> .....	IV-61
4.5.5.	<i>Test</i> .....	IV-66
4.6.	Kesimpulan.....	IV-66
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1.	Pendahuluan .....	V-1
5.2.	Data Hasil Percobaan/Penelitian .....	V-1
5.2.1.	Konfigurasi Percobaan.....	V-1
5.2.2.	Data Hasil Konfigurasi .....	V-2
5.3.	Analisis Hasil Penelitian.....	V-3
5.4.	Kesimpulan.....	V-8

BAB VI KESIMPULAN DAN SARAN .....	VI-1
6.1. Pendahuluan .....	VI-1
6.2. Kesimpulan.....	VI-1
6.3. Saran .....	VI-2
DAFTAR PUSTAKA .....	vi
LAMPIRAN.....	xi

## DAFTAR TABEL

	Halaman
<b>Tabel II-1.</b> Contoh NER pada Teks Berbahasa Indonesia .....	II-3
<b>Tabel II-2.</b> Contoh Proses Tokenisasi .....	II-7
<b>Tabel II-3.</b> Contoh Proses Stemming .....	II-8
<b>Tabel II-4.</b> Daftar Kelas Kata .....	II-9
<b>Tabel II-5.</b> Contoh Proses POS-Tagging .....	II-11
<b>Tabel II-6.</b> Daftar Word Features menurut (Bikel et al., 1999) .....	II-17
<b>Tabel II-7.</b> Daftar Word Features .....	II-18
<b>Tabel II-8.</b> Tabel Confusion Matrix .....	II-20
<b>Tabel II-9.</b> Tabel Multi-class Confusion Matrix .....	II-21
<b>Tabel IV-1.</b> Tabel Kebutuhan Fungsional .....	3
<b>Tabel IV-2.</b> Tabel Kebutuhan Non-Fungsional .....	4
<b>Tabel IV-3.</b> Contoh Label NER .....	5
<b>Tabel IV-4.</b> Contoh Data Latih .....	8
<b>Tabel IV-5.</b> Hasil Proses Tokenisasi .....	9
<b>Tabel IV-6.</b> Hasil Pemisahan Kata dengan Label NER .....	10
<b>Tabel IV-7.</b> Hasil Stemming .....	12
<b>Tabel IV-8.</b> Hasil Tahap POS-Tagging .....	13
<b>Tabel IV-9.</b> Data Hasil Pra-pemrosesan Teks dan POS-Tagging .....	14
<b>Tabel IV-10.</b> Nilai Probabilitas Transisi .....	16
<b>Tabel IV-11.</b> Nilai Probabilitas Emisi Bagian Pertama .....	18
<b>Tabel IV-12.</b> Nilai Probabilitas Emisi Bagian Kedua .....	18
<b>Tabel IV-13.</b> Nilai Probabilitas Emisi Bagian Ketiga .....	18
<b>Tabel IV-14.</b> Nilai Probabilitas Emisi Word Features Bagian Pertama .....	20
<b>Tabel IV-15.</b> Nilai Probabilitas Emisi Word Features Bagian Kedua .....	20
<b>Tabel IV-16.</b> Nilai Probabilitas Emisi Word Features Bagian Ketiga .....	20
<b>Tabel IV-17.</b> Contoh Data Uji .....	21
<b>Tabel IV-18.</b> Hasil Proses Tokenisasi .....	22
<b>Tabel IV-19.</b> Hasil Proses Stemming .....	23

<b>Tabel IV-20.</b> Hasil Tahap POS-Tagging.....	24
<b>Tabel IV-21.</b> Data Uji Hasil Pra-pemrosesan Teks dan POS-Tagging .....	25
<b>Tabel IV-22.</b> Hasil Tahap Pengujian.....	27
<b>Tabel IV-23.</b> Definisi Aktor .....	29
<b>Tabel IV-24.</b> Definisi Use Case .....	29
<b>Tabel IV-25.</b> Skenario Use Case Memasukkan Dokumen Data Latih.....	31
<b>Tabel IV-26.</b> Skenario Use Case Menghitung Nilai Parameter HMM .....	32
<b>Tabel IV-27.</b> Skenario Use Case Memasukkan Dokumen Data Uji.....	33
<b>Tabel IV-28.</b> Skenario Use Case Melakukan Pelabelan NER dan Menghitung Tingkat Performa Pelabelan .....	35
<b>Tabel IV-29.</b> Daftar Kelas-kelas yang dibutuhkan.....	51
<b>Tabel IV-30.</b> Skenario Pengujian Use Case Memilih Dokumen Data Latih .....	60
<b>Tabel IV-31.</b> Skenario Pengujian Use Case Menghitung Nilai Parameter HMM	60
<b>Tabel IV-32.</b> Skenario Pengujian Use Case Memilih Dokumen Data Uji.....	60
<b>Tabel IV-33.</b> Skenario Pengujian Use Case Melakukan Pelabelan NER dan Menghitung Tingkat Performa Pelabelan.....	61
<b>Tabel IV-34.</b> Hasil Pengujian Use Case Memilih Dokumen Data Latih .....	62
<b>Tabel IV-35.</b> Hasil Pengujian Use Case Menghitung dan Menyimpan Nilai Parameter HMM.....	63
<b>Tabel IV-36.</b> Hasil Pengujian Use Case Memilih Dokumen Data Uji.....	64
<b>Tabel IV-37.</b> Hasil Pengujian Use Case Melakukan Pelabelan NER dan Menghitung Tingkat Performa Pelabelan.....	65
<b>Tabel V-1.</b> Tabel Confusion Matrix dari Hasil Pengujian .....	V-2
<b>Tabel V-2.</b> Tabel Hasil Pengujian .....	V-3



## DAFTAR GAMBAR

	Halaman
<b>Gambar III-1.</b> Diagram Tahapan Sistem NER menggunakan HMM dan POS- Tagging .....	III-4
<b>Gambar IV-1.</b> Gambaran Data Latih .....	IV-6
<b>Gambar IV-2.</b> Gambaran Data Uji .....	IV-7
<b>Gambar IV-3.</b> Contoh Perhitungan Probabilitas Transisi .....	IV-16
<b>Gambar IV-4.</b> Contoh Perhitungan Probabilitas Emisi .....	IV-17
<b>Gambar IV-5.</b> Contoh Perhitungan Probabilitas Emisi Word Features .....	IV-19
<b>Gambar IV-6.</b> Contoh Perhitungan Algoritma Viterbi.....	IV-26
<b>Gambar IV-7.</b> Contoh Perhitungan Algoritma Viterbi Menggunakan Probabilitas Emisi Word Features .....	IV-26
<b>Gambar IV-8.</b> Diagram Use Case.....	IV-28
<b>Gambar IV-9.</b> Desain Antarmuka Proses Pelatihan .....	IV-39
<b>Gambar IV-10.</b> Desain Antarmuka Proses Pengujian .....	IV-39
<b>Gambar IV-11.</b> Activity Diagram Memasukkan Dokumen Data Latih .....	IV-40
<b>Gambar IV-12.</b> Activity Diagram Menghitung Nilai Parameter HMM.....	IV-41
<b>Gambar IV-13.</b> Activity Diagram Memasukkan Dokumen Data Uji .....	IV-41
<b>Gambar IV-14.</b> Activity Diagram Melakukan Pelabelan NER dan Menghitung Tingkat Performa Pelabelan .....	IV-42
<b>Gambar IV-15.</b> Sequence Diagram Memasukkan Dokumen Data Latih.....	IV-43
<b>Gambar IV-16.</b> Sequence Diagram Menghitung dan Menyimpan Nilai Parameter HMM .....	IV-44
<b>Gambar IV-17.</b> Sequence Diagram Memasukkan Dokumen Data Uji .....	IV-45

<b>Gambar IV-18.</b> Sequence Diagram Melakukan Pelabelan NER dan Menghitung Tingkat Performa Pelabelan .....	IV-46
<b>Gambar IV-19.</b> Sequence Diagram Preprocessing.....	IV-47
<b>Gambar IV-20.</b> Sequence Diagram POS-Tagging .....	IV-47
<b>Gambar IV-21.</b> Sequence Diagram HMM Training .....	IV-48
<b>Gambar IV-22.</b> Sequence Diagram HMM Classifier.....	IV-49
<b>Gambar IV-23.</b> Sequence Diagram Evaluation.....	IV-49
<b>Gambar IV-24.</b> Class Diagram Perangkat Lunak.....	IV-56
<b>Gambar IV-25.</b> Tampilan Antarmuka Perangkat Lunak Fase Pelatihan.....	IV-57
<b>Gambar IV-26.</b> Tampilan Antarmuka Perangkat Lunak Fase Pengujian.....	IV-57
<b>Gambar V-1.</b> Perbandingan Nilai F-Measure pada Label NER .....	V-3
<b>Gambar V-2.</b> Perbandingan Jumlah Kata Diketahui dan Kata Tidak Diketahui pada Data Uji.....	V-4
<b>Gambar V-3.</b> Perbandingan Nilai Recall dan Precision pada Label NER.....	V-5
<b>Gambar V-5.</b> Perbandingan Nilai F-Measure dengan Sistem NER Tanpa POS-Tagging.....	V-7

## DAFTAR LAMPIRAN

	Halaman
1	Contoh Data Latih.....L-2
2	Contoh Data Uji .....L-9
3	<i>Source Code</i> Program .....L-16

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab ini akan membahas mengenai penjelasan umum keseluruhan penelitian, diantaranya adalah latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup, sistematika penulisan dan kesimpulan.

### **1.2 Latar Belakang Masalah**

*Named-Entity Recognition* (NER) adalah salah satu topik penelitian pada bidang Pemrosesan Bahasa Alami (PBA) yang melibatkan pemrosesan teks dari dokumen-dokumen yang telah terstruktur maupun tidak terstruktur, lalu mengidentifikasi entitas kata yang merujuk pada nama orang, nama tempat, nama organisasi dan satuan waktu (Mansouri dan Affendey, 2008). Menurut Fachri (2014), NER melibatkan dua tugas, yang pertama adalah identifikasi entitas kata yang tepat di dalam teks, lalu yang kedua adalah mengklasifikasikan entitas-entitas ini menjadi beberapa kategori entitas kata yaitu *Person*, *Location*, *Organization*, dan *Time*.

Kemampuan untuk mengenali entitas kata lalu mengklasifikasikannya merupakan bagian penting dari *Named-Entity Recognition* (NER). Kemampuan tersebut bergantung pada cara mengenali (*recognition*) dan adanya aturan klasifikasi (*classification*) yang dipicu oleh ciri-ciri tertentu pada kata tersebut. Pada era awal sebagian besar masih didasar pada aturan-aturan yang sederhana, dan

yang paling baru sekarang adalah menggunakan pembelajaran terarah (*Supervised Learning*) (Nadeau dan Sekine, 2009).

Menurut laporan studi yang disusun oleh Sharnagat (2014), terdapat lima metode *Supervised Learning* yang sering digunakan untuk memecahkan masalah *Named-Entity Recognition* (NER), diantaranya adalah *Hidden Markov Models* (HMM), *Decision Trees*, *Maximum Entropy* (ME), *Support Vector Machines* (SVM), dan *Conditional Random Fields* (CRF).

*Hidden Markov Model* (HMM) menjadi metode yang digunakan dalam penelitian ini dikarenakan metode ini memiliki kelebihan yaitu kecepatan pelatihan metode ini lebih baik dari metode lainnya, dan tepat digunakan dengan penelitian yang membutuhkan data pelatihan berupa data teks (Sharma dan Lehal, 2011). Selain itu, HMM menjadi metode yang populer dalam menyelesaikan masalah *Named-Entity Recognition* (NER) dikarenakan efisiensi waktu yang sangat baik dalam melakukan tahap pengujian berkat penggunaan algoritma *Viterbi* (Ekbal dan Bandyopadhyay, 2007). Terakhir, HMM merupakan metode yang mudah untuk dipahami, mudah untuk diimplementasikan, dan dapat dianalisis dengan baik (Morwal et al., 2012).

Penggunaan metode *Hidden Markov Model* (HMM) dalam menyelesaikan masalah *Named-Entity Recognition* (NER) telah dilakukan oleh Ekbal dan Bandyopadhyay (2007), Zia dan Sharifi (2019), dan (Sari et al., 2019). Pada penelitian yang dilakukan oleh Ekbal dan Bandyopadhyay (2007), mereka menyelesaikan masalah NER pada teks berbahasa Bengali dan Hindi, dan berhasil mendapatkan nilai *F-Measure* sebesar 84,50% untuk bahasa Bengali, dan 78,35%

untuk bahasa Hindi. Lalu pada penelitian Zia dan Sharifi (2019), mereka berhasil menyelesaikan masalah NER pada teks berbahasa Dari (bahasa resmi Afghanistan) dengan nilai *F-Measure* sebesar 93.7%. Terakhir, pada penelitian yang dilakukan oleh Sari et al. (2019), mereka berhasil menyelesaikan masalah NER pada teks berbahasa Indonesia dengan nilai *F-Measure* sebesar 86%.

Meskipun menurut Blunsom (2004), *Hidden Markov Model* (HMM) merupakan metode yang populer dan memiliki tingkat keberhasilan yang tinggi, sama seperti metode-metode *Supervised Learning* lainnya, tingkat keberhasilan metode HMM bergantung dari *data training* yang digunakan, jika pada saat tahap pengujian, ternyata data yang diuji tidak terdapat pada *data training*, maka dapat menyebabkan tidak tepatnya metode HMM dalam melakukan klasifikasi data.

Pada penelitian yang dilakukan oleh Bikel et al. (1999), keterbatasan dari metode *Hidden Markov Model* (HMM) ini dicoba diatasi dengan penambahan algoritma baru, baik pada tahap pelatihan maupun tahap pengujiannya, yaitu dengan menggunakan *Word Features* sebagai kata ganti dari kata yang ternyata tidak ada dalam *data training*, tetapi menurut Bikel, hasil yang didapat tidak terlalu signifikan, yaitu dari *F-measure* awal sebesar 94.07% menjadi 94.92% pada saat mengimplementasikan *Word features*.

Pada penelitian yang dilakukan oleh Setiyoaji et al. (2017), mereka mencoba untuk menyelesaikan masalah *Named-Entity Recognition* (NER) pada teks berbahasa Indonesia yang memiliki kaitan dengan tanaman obat, dengan menggunakan metode *Hidden Markov Model* (HMM) dan Algoritma *Viterbi*. Hasil pengujian validasi pada penelitian ini adalah 79.02% untuk nilai *F-measure*. Pada

penelitian ini, Setiyoaji et al. (2017) memberikan saran yaitu diperlukannya proses *POS-Tagging* pada setiap kata, yang selanjutnya digunakan untuk membantu metode HMM dalam mengklasifikasikan kategori entitas kata.

*POS-Tagging* itu sendiri menurut Rashel et al. (2014) adalah proses untuk memberikan label kelas kata pada suatu kata secara otomatis, beberapa contoh dari label kelas kata pada bahasa Indonesia yaitu kata benda, kata sifat, dan kata kerja. Berdasarkan buku yang telah ditulis oleh Jurafsky dan Martin (2000), *POS-Tagging* merupakan proses yang penting, karena dapat menunjukkan hubungan sebuah kata dengan kata lainnya pada suatu kalimat, hubungan antar kata ini menjadi salah satu fitur penting yang selanjutnya bisa digunakan pada saat menyelesaikan masalah *Named-Entity Recognition* (NER).

Pada penelitian yang dilakukan oleh Rifani et al. (2019) tentang penyelesaian masalah *Named-Entity Recognition* (NER) menggunakan *Multinomial Naïve Bayes Classifier*, mereka menggunakan tiga macam *Feature Extraction* untuk membantu metode Naïve Bayes, yaitu *POS-Tagging*, *Unigram*, dan *Laplacian Smoothing*. Dari hasil penelitian tersebut, diberikan kesimpulan bahwa fitur *POS-Tagging* merupakan fitur yang paling berpengaruh, karena dapat meningkatkan hasil akurasi yang lebih baik dibandingkan dengan fitur lainnya, yaitu meningkatkan nilai *F-Measure* dari 61% menjadi 80%.

Berdasarkan penelitian-penelitian yang telah dijelaskan diatas, khususnya dari penelitian Setiyoaji et al. (2017), maka diperlukannya penelitian lebih lanjut pada metode *Hidden Markov Model* (HMM) sehingga dapat menyelesaikan masalah *Named-Entity Recognition* (NER) dengan baik, khususnya pada teks



berbahasa Indonesia. Pada penelitian ini, hasil dari proses *POS-Tagging* akan digunakan pada saat menangani kata yang tidak terdapat pada *data training*. Cara untuk menggunakan hasil dari *POS-Tagging* sehingga dapat membantu metode HMM dalam menyelesaikan masalah NER akan dilakukan berdasarkan penelitian dari Bikel et al. (1999), yaitu dengan menggunakan *Word Features*. *Word Features* akan diambil berdasarkan kelas kata dari hasil *POS-Tagging*. Dengan demikian, pada penelitian ini akan dilakukan penyelesaian masalah *Named-Entity Recognition* (NER) pada teks Berbahasa Indonesia dengan menggunakan metode *Hidden Markov Model* (HMM) dan *POS-Tagging*.

### 1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah bagaimana tingkat performa dari metode *Hidden Markov Model* (HMM) dan *POS-Tagging* dalam melakukan *Named-Entity Recognition* (NER), atau dengan kata lain yaitu mengklasifikasikan entitas suatu kata.

Untuk menyelesaikan permasalahan yang disebutkan diatas, maka dalam penelitian ini dibagi dalam beberapa *research question* antara lain:

1. Bagaimana arsitektur sistem *Named-Entity Recognition* (NER) menggunakan metode *Hidden Markov Model* (HMM) dan *POS-Tagging*?
2. Bagaimana mekanisme pengembangan perangkat lunak yang dapat mengklasifikasikan entitas suatu kata menggunakan metode *Hidden Markov Model* (HMM) dan *POS-Tagging*?
3. Bagaimana tingkat performa dari metode *Hidden Markov Model* (HMM) dan *POS-Tagging* dalam mengklasifikasikan entitas suatu kata?

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Menghasilkan arsitektur sistem *Named-Entity Recognition* (NER) menggunakan metode *Hidden Markov Model* (HMM) dan *POS-Tagging*.
2. Mengetahui mekanisme pengembangan perangkat lunak yang dapat mengklasifikasikan entitas suatu kata menggunakan metode *Hidden Markov Model* (HMM) dan *POS-Tagging*.
3. Mengetahui tingkat performa dari metode *Hidden Markov Model* (HMM) dan *POS-Tagging* dalam mengklasifikasikan entitas suatu kata. Tingkat performa metode diukur berdasarkan nilai *recall*, *precision*, dan *f-measure* dalam melakukan *Named-Entity Recognition* (NER).

#### **1.5 Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Perangkat lunak dari hasil penelitian dapat digunakan untuk mengklasifikasikan entitas suatu kata pada teks berbahasa Indonesia.
2. Dapat menjadi salah satu referensi dalam penelitian di bidang Pemrosesan Bahasa Alami (PBA), khususnya pada bidang *Named-Entity Recognition* (NER).

#### **1.6 Batasan Masalah**

Batasan permasalahan yang akan diambil dalam penelitian ini adalah sebagai berikut:

1. Data latih dan data uji berupa kalimat berbahasa Indonesia.

2. Data latih dan data uji yang akan digunakan berupa dokumen dengan format TXT.
3. Kategori entitas nama yang diklasifikasikan yaitu nama orang (PER), nama lokasi (LOC), nama organisasi (ORG), satuan waktu (TIME), dan selain kategori diatas (OTH).

### **1.7 Sistematika Penulisan**

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

#### **BAB I. PENDAHULUAN**

Bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup, sistematika penulisan dan kesimpulan.

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

Bab ini akan menjelaskan dasar-dasar teori yang digunakan dalam penelitian. Pembahasan dalam bab ini meliputi penjelasan mengenai Pemrosesan Bahasa Alami, *Named-Entity Recognition* (NER), pra-pemrosesan teks, proses *POS-Tagging*, metode *Hidden Markov Model* (HMM), algoritma *Viterbi*, *Word Features*, *Confusion Matrix*, *Rational Unified Process* (RUP), *Unified Modelling Language* (UML), dan menampilkan hasil penelitian terdahulu yang relevan dengan penelitian ini.

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

Bab ini akan membahas tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci

dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

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

Bab ini akan membahas tahapan-tahapan yang dilakukan dalam proses pengembangan perangkat lunak berdasarkan metode *Rational Unified Process* (RUP).

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

Bab ini akan membahas hasil pengujian berdasarkan rencana pengujian yang telah ditentukan. Analisis hasil penelitian akan diberikan sebagai dasar dari kesimpulan yang diambil pada penelitian ini.

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

Bab ini berisikan kesimpulan dan saran dari penelitian ini. Kesimpulan dan saran diambil dari semua pembahasan yang dilakukan pada bab sebelumnya.

### **1.8 Kesimpulan**

Pada penelitian ini, akan dijelaskan bagaimana cara menerapkan metode *Hidden Markov Model* (HMM) dalam melakukan *Named-Entity Recognition* (NER). Pada penelitian ini juga akan melakukan proses *POS-Tagging*, lalu menerapkannya kedalam proses HMM dengan tujuan untuk membantu metode HMM dalam melakukan NER. Selanjutnya, dilakukan pengujian untuk melihat tingkat performa hasil klasifikasi yang dicapai dalam melakukan penelitian ini.

## DAFTAR PUSTAKA

- Bikel, D. M., Schwartz, R., & Weischedel, R. M. (1999). Algorithm that learns what's in a name. *Machine Learning*, 34(1), 211–231.  
<https://doi.org/10.1023/A:1007558221122>
- Blunsom, P. (2004). Hidden Markov Models. *Lecture notes, August, 15*, 48.  
[https://doi.org/10.1007/978-1-61779-400-1\\_22](https://doi.org/10.1007/978-1-61779-400-1_22)
- Dirgantara, M. Y. S., Fauzi, M. A., & Perdana, R. S. (2018). Penerapan Named Entity Recognition Untuk Mengenal Fitur Produk Pada E-Commerce Menggunakan Rule Template Dan Hidden Markov Model. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(10), 3912–3920.
- Ekbal, A., & Bandyopadhyay, S. (2007). A Hidden Markov Model based named entity recognition system: Bengali and Hindi as case studies. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4815 LNCS, 545–552.  
[https://doi.org/10.1007/978-3-540-77046-6\\_67](https://doi.org/10.1007/978-3-540-77046-6_67)
- Fachri, M. (2014). *Named Entity Recognition For Indonesian Text Using Hidden Markov Model*. Universitas Gadjah Mada.
- Haulrich, M. (2009). *Different Approaches to Unknown Words in a Hidden Markov Model Part-of-Speech Tagger Tagging with HMMs*. 1–8.  
<http://stp.lingfil.uu.se/~nivre/statmet/haulrich.pdf>
- Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and*

- Speech Recognition* (Third Edit). <https://doi.org/10.1515/zfsw.2002.21.1.134>
- Kale, S., & Govilkar, S. (2017). Survey of Named Entity Recognition Techniques for Various Indian Regional Languages. *International Journal of Computer Applications*, 164(4), 37–43. <https://doi.org/10.5120/ijca2017913621>
- Konkol, M. (2012). *Named Entity Recognition PhD Study Report*. <https://www.kiv.zcu.cz/site/documents/verejne/vyzkum/publikace/technicke-zpravy/2012/tr-2012-04.pdf>
- Mansouri, A., & Affendey, L. (2008). Named entity recognition approaches. *Ijcsns*, 8(2), 339–344. <https://doi.org/10.1075/li.30.1.03nad>
- Morwal, S., Jahan, N., & Chopra, D. (2012). Named Entity Recognition using Hidden Markov Model (HMM). *International Journal on Natural Language Computing*, 1(4), 15–23. <https://doi.org/10.5121/ijnlc.2012.1402>
- Nadeau, D., & Sekine, S. (2009). A survey of named entity recognition and classification. *Lingvisticae InvestigationesLingvisticae InvestigationesLingvisticae Investigationes. International Journal of Linguistics and Language Resources*, 30(1), 3–26. <https://doi.org/10.1075/li.30.1.03nad>
- Nazief, B., Adriani, M., Asian, J., Tahaghoghi, S. M. M., & Williams, H. E. (2005). Stemming Indonesian. *Conferences in Research and Practice in Information Technology Series*, 38(January), 307–314. <https://doi.org/10.1145/1316457.1316459>
- Poria, S. B. S., Gelbukh, A., Hussain, A., & Das, D. (2013). Enhanced SenticNet with Affective Labels for Concept-based Opinion Mining. *IEEE Intelligent*

*System.*

- Rashel, F., Luthfi, A., Dinakaramani, A., & Manurung, R. (2014). Building an Indonesian rule-based part-of-speech tagger. *Proceedings of the International Conference on Asian Language Processing 2014, IALP 2014*, 70–73. <https://doi.org/10.1109/IALP.2014.6973521>
- Rifani, R., Bijaksana, M. A., & Asror, I. (2019). Named Entity Recognition for an Indonesian Based Language Tweet using Multinomial Naive Bayes Classifier. *Indonesia Journal of Computing*, 4(2), 119–126. <https://doi.org/10.21108/indojc.2019.4.2.330>
- Robbins, H. A. (2017). *Unified Modeling Language (UML): Analysis, System, and Method* (Patent No. 9672010). <https://patents.justia.com/patent/9672010>
- Sang, E. F. T. K., & Meulder, F. De. (2003). *Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition*. <http://arxiv.org/abs/cs/0306050>
- Sari, W. P., Bijaksana, M. A., & Huda, A. F. (2019). Indexing name in hadith translation using hidden markov model (HMM). *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*, 1–5. <https://doi.org/10.1109/ICoICT.2019.8835296>
- Setiyoadji, 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*, 1(12), 1858–1864. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/673>



- Sharma, S. K., & Lehal, G. S. (2011). Using Hidden Markov Model to improve the accuracy of Punjabi POS tagger. *Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011*, 2, 697–701. <https://doi.org/10.1109/CSAE.2011.5952600>
- Sharnagat, R. (2014). *Named Entity Recognition: A Literature Survey*. [http://www.cfilt.iitb.ac.in/resources/surveys/NER-Literature Survey-2012-Satpreet.pdf](http://www.cfilt.iitb.ac.in/resources/surveys/NER-Literature%20Survey-2012-Satpreet.pdf)
- Sukanto, & Shalahuddin, M. (2013). *Analisa dan Desain Sistem Informasi*. Andi Offset.
- Syaifudin, Y. (2016). *Quotations Identification From Indonesian Online News Using Rule-Based Method*. Universitas Gadjah Mada.
- Tala, F. Z. (2003). *A Study Of Stemming Effects On Information Retrieval in Bahasa Indonesia*.
- Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson Addison-Wesley.
- Weiss, S. M., Indurkha, N., Zhang, T., & Damerau, F. J. (2005). *Text Mining: Predictive Methods fo Analyzing Unstructured Information*. Springer.
- Wibisono, Y. (2008). *Penggunaan Hidden Markov Model untuk Kompresi Kalimat*. 5–17. <http://digilib.itb.ac.id/files/disk1/627/jbptitbpp-gdl-yudiwibiso-31314-1-2008ts-r.pdf>
- Wicaksono, A. F., & Purwarianti, A. (2010). *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*.
- Wu, S., Fang, Z., & Tang, J. (2012). Accurate Product Name Recognition from User

Generated Content. *2012 IEEE 12th International Conference on Data Mining Workshops*, 874–877. <https://doi.org/10.1109/ICDMW.2012.129>

Zia, G. A. J., & Sharifi, A. Z. (2019). *HMM-Based Dari Named Entity Recognition for Information Extraction. MI*, 73–81. <https://doi.org/10.5121/csit.2019.90706>