

**NAMED ENTITY RECOGNITION MENGGUNAKAN HIDDEN
MARKOV MODEL DAN ALGORITMA VITERBI PADA TEKS
TANAMAN OBAT**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:
Agung Setiyoaji
NIM: 135150201111114



PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2017

LEMBAR PENGESAHAN

NAMED ENTITY RECOGNITION MENGGUNAKAN HIDDEN MARKOV MODEL DAN
ALGORITMA VITERBI PADA TEKS TANAMAN OBAT

SKRIPSI

KEMINATAN TEKNIK INFORMATIKA

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
AGUNG SETIYOAJI
NIM: 135150201111114

Skripsi ini telah diuji dan dinyatakan lulus pada
7 Juli 2017

Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I

Dosen Pembimbing II

Lailil Muflikhah, S.Kom, M.Sc
NIP. 19741113 200501 2 001

M. Ali Fauzi, S.Kom, M.Kom
NIK. 201502 890101 1 001

Mengetahui
Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T, M.T, Ph.D
NIP. 19710518 200312 1 001

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 7 Juli 2017

Agung Setiyoadi

NIM: 135150201111114

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, hingga saya dapat menyelesaikan skripsi dengan judul “Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat”

Skripsi merupakan salah satu syarat untuk menyelesaikan studi di program studi informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, Malang. Tidak dapat dipungkiri bahwa tidak mungkin penulis dapat menyelesaikan skripsi ini tanpa bantuan dan dukungan dari banyak pihak. Dalam kesempatan ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Orang tua dan saudara kandung saya yang telah menyemangati dan memberi dukungan moral dan materi.
2. Ibu Lailil Muflikhah, S.Kom, M.Sc selaku pembimbing 1, terima kasih atas sara, ilmu, dan waktu yang telah diberikan kepada penulis dalam pelaksanaan skripsi.
3. Bapak M. Ali Fauzi, S.Kom, M.Kom selaku pembimbing 2, terima kasih atas saran dan ilmu yang telah diberikan kepada penulis dalam pelaksanaan skripsi.
4. Bapak Wayan Firdaus Mahmudy, S.Si, M.T., Ph.D selaku Dekan Fakultas Ilmu Komputer Universitas Brawijaya Malang.
5. Bapak Tri Astoto Kurniawan, S.T, M.T, Ph.D selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Brawijaya Malang.
6. Sahabat-sahabat yang selalu memberi dukungan dan motivasi untuk pelaksanaan dan pengerjaan skripsi.

7. Teman-teman Informatika/Ilmu Komputer Universitas Brawijaya dan seluruh pihak yang telah membantu kelancaran pelaksanaan dan penyusunan skripsi.

Penulis menyadari skripsi ini jauh dari kata sempurna. Oleh karena itu, penulis dengan senang hati dan terbuka menerima kritik dan saran yang bersifat membangun.

Malang, 7 Juli 2017

setiyoajiagung@gmail.com

ABSTRAK

Media untuk menyampaikan informasi dapat melalui televisi, radio, social media, dan website. Website merupakan karya tulis dari seseorang terletak pada sebuah domain yang mengandung informasi. Perkembangan website semakin banyak dan informasi semakin tidak terbandung sehingga muncul permasalahan sulitnya mencari informasi yang sesuai dengan kebutuhan pengguna internet, sehingga dibutuhkan klasifikasi dan ekstraksi informasi untuk informasi yang ada dalam website. *Named Entity Recognition* (NER) yang merupakan turunan dari ekstraksi informasi, bertujuan untuk memudahkan mencari informasi dengan cara pemberian nama entitas pada setiap kata dalam sebuah teks. Pada penelitian ini akan dilakukan pengenalan empat entitas yaitu NAMA, TEMPAT, ZAT, dan KEGUNAAN dari teks tanaman obat. Algoritma pada penelitian ini menggunakan *Hidden Markov Model* (HMM) dan algoritma Viterbi. Secara keseluruhan pengenalan entitas menghasilkan nilai yang paling rendah dengan *f-measure* 0.41, dan yang paling tinggi dengan *f-measure* 0.72.

Kata kunci: *Hidden Markov Model, Named Entity Recognition, tanaman obat, ekstraksi informasi*

ABSTRACT

Media to convey information can be through television, radio, social media, and website. Website is a work of someone located in a domain that contains information. The development of websites more and more information is not unstoppable so that the problem arises difficult to find information in accordance with the needs of Internet users, so that the required classification and extraction of information for information on the website. Named Entity Recognition (NER) which derives from the extraction of information, NER aims to facilitate the search for information by naming entities on each word in a text. In this research will be done the introduction of four entities namely the NAME, PLACE, SUBSTANCE, and FUNCTION of the text on medicinal plants. The algorithm used Hidden Markov Model (HMM) and Viterbi algorithm. Overall entity recognition count the lowest value with f-measure 0.41, and the highest with f-measure 0.72.

Keyword: *Hidden Markov Model, Named Entity Recognition, tanaman obat, ekstraksi informasi*

DAFTAR ISI

NAMED ENTITY RECOGNITION MENGGUNAKAN HIDDEN MARKOV MODEL DAN ALGORITMA VITERBI PADA TEKS TANAMAN OBAT	1
LEMBAR PENGESAHAN	2
PERNYATAAN ORISINALITAS	3
KATA PENGANTAR	4
ABSTRAK	6
ABSTRACT	7
DAFTAR ISI	8
DAFTAR TABEL	11
DAFTAR GAMBAR	12
DAFTAR LAMPIRAN	13
BAB 1 PENDAHULUAN	Error!
Bookmark not defined.	
1.1 Latar belakang	Error! Bookmark not defined.
1.2 Rumusan masalah	Error! Bookmark not defined.
1.3 Tujuan	Error! Bookmark not defined.
1.4 Manfaat	Error! Bookmark not defined.
1.5 Batasan masalah	Error! Bookmark not defined.
1.6 Sistematika pembahasan	Error! Bookmark not defined.
BAB 2 LANDASAN KEPUSTAKAAN	Error!
Bookmark not defined.	
2.1 Kajian pustaka	Error! Bookmark not defined.
2.2 Tanaman obat	Error! Bookmark not defined.
2.3 Ekstraksi informasi	Error! Bookmark not defined.
2.4 Preprosesing	Error! Bookmark not defined.
2.4.1 Tokenisasi	Error! Bookmark not defined.
2.4.2 Filtering	Error! Bookmark not defined.
2.4.3 Stemming	Error! Bookmark not defined.

2.5 Named Entity Recognition (NER)Error! Bookmark not defined.
2.6 Hidden Markov Model (HMM)Error! Bookmark not defined.
2.7 Algoritma ViterbiError! Bookmark not defined.
BAB 3 METODOLOGIError!
Bookmark not defined.	
3.1 Studi LiteraturError! Bookmark not defined.
3.2 PerancanganError! Bookmark not defined.
3.2.1 Kebutuhan dataError! Bookmark not defined.
3.3 ImplementasiError! Bookmark not defined.
3.4 Pengujian dan analisisError! Bookmark not defined.
3.5 KesimpulanError! Bookmark not defined.
BAB 4 PERANCANGANError!
Bookmark not defined.	
4.1 Diskripsi umum sistemError! Bookmark not defined.
4.2 Kebutuhan sistemError! Bookmark not defined.
4.2.1 Kebutuhan perangkat kerasError! Bookmark not defined.
4.2.2 Kebutuhan perangkat lunakError! Bookmark not defined.
4.3 Diagram alir sistemError! Bookmark not defined.
4.4 Contoh penghitungan manualError! Bookmark not defined.
4.4.1 Penghitungan probabilitas awalError! Bookmark not defined.
4.4.2 Penghitungan probabilitas transisiError! Bookmark not defined.
4.4.3 Penghitungan probabilitas emisiError! Bookmark not defined.
4.4.4 Penghitungan algoritme viterbiError! Bookmark not defined.
4.5 Perancangan antarmukaError! Bookmark not defined.
4.6 Perancangan pengujianError! Bookmark not defined.
4.7 Penarikan kesimpulanError! Bookmark not defined.
BAB 5 IMPLEMENTASIError!
Bookmark not defined.	
5.1 Spesifikasi sistemError! Bookmark not defined.
5.1.1 Spesifikasi perangkat kerasError! Bookmark not defined.

5.1.2 Spesifikasi perangkat lunak	Error! Bookmark not defined.
5.2 Implementasi program	Error! Bookmark not defined.
5.2.1 Implementasi probabilitas awal	Error! Bookmark not defined.
5.2.2 Implementasi probabilitas transisi	Error! Bookmark not defined.
5.2.3 Implementasi probabilitas emisi	Error! Bookmark not defined.
5.2.4 Implementasi algoritma viterbi	Error! Bookmark not defined.
5.3 Implementasi antarmuka	Error! Bookmark not defined.
BAB 6 PENGUJIAN DAN ANALISA HASIL	Error!
Bookmark not defined.		
6.1 Skenario pengujian	Error! Bookmark not defined.
6.1.1 Pengujian validasi	Error! Bookmark not defined.
6.1.2 Pengujian dan analisis 30 data uji	Error! Bookmark not defined.
6.1.3 Pengujian dan analisis 30 data uji berbeda	Error! Bookmark not defined.
6.1.4 Pengujian dan analisis penambahan pos-tagging	Error! Bookmark not defined.
6.2 Analisis hasil pengujian	Error! Bookmark not defined.
BAB 7 PENUTUP	Error!
Bookmark not defined.		
7.1 Kesimpulan	Error! Bookmark not defined.
7.2 Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA	Error!
Bookmark not defined.		
LAMPIRAN A	Error!
Bookmark not defined.		
A.1 Sampel data penelitian	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 2.1 Contoh Tokenisasi	7
Tabel 2.2 Contoh hasil <i>filtering</i>	8
Tabel 2.3 Contoh Stemming.....	8
Tabel 3.1 Data penelitian	16
Tabel 4.1 Contoh data latih.....	26
Tabel 4.2 Penghitungan probabilitas transisi.....	30
Tabel 4.3 Penghitungan probabilitas emisi.....	31
Tabel 4.4 Contoh data uji	33
Tabel 4.5 Penghitungan algoritme viterbi.....	33
Tabel 5.1 Implementasi probabilitas awal	37
Tabel 5.2 Implementasi probabilitas transisi	38
Tabel 5.3 Implementasi probabilitas emisi	40
Tabel 5.4 Implementasi algoritma viterbi.....	42
Tabel 6.1 Pengujian Validasi.....	48
Tabel 6.2 Pengujian 30 data uji	49
Tabel 6.3 Pengujian perbedaan data	49
Tabel 6.4 Pengujian pos-tagging	50

DAFTAR GAMBAR

Gambar 2.1 Bentuk umum imbuhan	9
Gambar 2.2 Contoh tipe kata	10
Gambar 2.3 Contoh <i>POS-Tagging</i>	10
Gambar 2.4 Markov Chain Cuaca.....	12
Gambar 3.1 Tahapan penelitian.....	15
Gambar 4.1 Deskripsi umum sistem	18
Gambar 4.2 Alir sistem keseluruhan	20
Gambar 4.3 Diagram alir HMM.....	21
Gambar 4.4 Diagram alir probabilitas awal	23
Gambar 4.5 Diagram alir probabilitas transisi	24
Gambar 4.6 Diagram alir probabilitas emisi	25
Gambar 4.7 Diagram alir algoritme viterbi	26
Gambar 5.1 Tampilan menu	46
Gambar 5.2 Tampilan input data	46
Gambar 5.3 Tampilan hasil pengenalan.....	47

DAFTAR LAMPIRAN

Lampiran 1. Sampel data penelitian	55
--	----