

SKRIPSI

MORPHOLOGICAL PARSER UNTUK TEKS DALAM
BAHASA INDONESIA



Andreas Novian Dwi Triastanto

NPM: 2013730038

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2017

UNDERGRADUATE THESIS

**MORPHOLOGICAL PARSER FOR TEXTS IN BAHASA
INDONESIA**



Andreas Novian Dwi Triastanto

NPM: 2013730038

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2017**

LEMBAR PENGESAHAN



**MORPHOLOGICAL PARSER UNTUK TEKS DALAM
BAHASA INDONESIA**

Andreas Novian Dwi Triastanto

NPM: 2013730038

Bandung, 12 Desember 2017

Menyetujui,

Pembimbing

Dott. Thomas Anung Basuki

Ketua Tim Penguji

Claudio Franciscus, M.T.

Anggota Tim Penguji

Pascal Alfadian, M.Comp.

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng



PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

MORPHOLOGICAL PARSER UNTUK TEKS DALAM BAHASA INDONESIA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 12 Desember 2017



Andreas Novian Dwi Triastanto
NPM: 2013730038

ABSTRAK

Morphological parser adalah salah satu alat dalam pengolahan bahasa alami yang berfungsi untuk membagi sebuah kata menjadi komponen-komponen penyusunnya, seperti kata dasar, awalan, akhiran, serta dapat mengenali jika kata tersebut merupakan kata ulang maupun kata majemuk. Proses ini dapat membantu mengurangi ambiguitas selama proses mengetahui makna suatu kalimat. Perangkat lunak yang dibuat pada penelitian ini dapat mengenali kata dalam bahasa Indonesia yang dibentuk dari proses morfologi berupa afiksasi, reduplikasi, dan komposisi. Leksikon dibutuhkan dalam perangkat lunak untuk menyimpan semua kata dasar dan kata turunan yang valid dalam bahasa Indonesia. Hasil dari penelitian ini menunjukkan bahwa perangkat lunak telah mampu untuk melakukan morphological parsing pada kata dalam bahasa Indonesia dan mampu menghasilkan semua kemungkinan parsing dari sebuah kata. Perangkat lunak dapat melakukan proses parsing untuk teks yang terdiri dari 100 kata dalam waktu rata-rata kurang dari 100 milidetik.

Kata-kata kunci: morphological parser, morphological parsing, pengolahan bahasa alami, bahasa Indonesia, leksikon

ABSTRACT

Morphological parser is one of the tools used in natural language processing to decompose a word into its components, such as root, prefix, suffix, and can also recognize if the word is a reduplication or compound word. This process can help reducing the ambiguity during the process of understanding the meaning of a sentence. The software built in this research can recognize a word in bahasa Indonesia formed from the morphological process of affixation, reduplication, and composition. Lexicon is needed in the software to keep all the valid root words and their derivatives in bahasa Indonesia. The result of this research shows that the software is capable to do morphological parsing on words in bahasa Indonesia and capable of producing all possible parsing of a word. The software is capable of doing parsing to text consisting of 100 words in less than 100 milliseconds.

Keywords: morphological parser, morphological parsing, natural language processing, bahasa Indonesia, lexicon

untuk keluarga, bangsa, dan almamater

KATA PENGANTAR

Puji syukur penulis sampaikan kepada Tuhan Yesus Kristus atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul '*Morphological Parser* untuk Teks dalam Bahasa Indonesia'. Penulis juga ingin mengucapkan terima kasih kepada pihak-pihak yang telah membantu dan memberikan dukungan kepada penulis, yaitu:

1. Bapak Thomas Anung Basuki selaku dosen pembimbing yang telah meluangkan waktunya untuk mengarahkan dan memberikan masukan selama proses penulisan karya tulis ini.
2. Bapak Agustinus Triyono dan Ibu Budi Astuti Mulatsih yang telah memberikan dukungan jasmani dan rohani kepada penulis.
3. Bapak Claudio Franciscus dan Bapak Pascal Alfadian Nugroho selaku tim penguji yang telah memberikan kritik dan saran yang berharga pada karya tulis ini.
4. Pengurus Persatuan Mahasiswa Universitas Katolik Parahyangan periode 2014/2015 dan 2015/2016 yang telah memberikan banyak pengalaman dan motivasi.
5. Teman-teman di program studi Teknik Informatika Unpar yang bersama-sama berjuang menyelesaikan setiap tugas dan ujian selama masa kuliah.

Penulis berharap melalui penelitian yang dilakukan pada skripsi ini dapat membantu orang-orang yang juga ingin mengembangkan sistem pengolahan bahasa alami khususnya untuk bahasa Indonesia. Akhir kata penulis mengharapkan kritik dan saran dari para pembaca supaya tulisan ini menjadi lebih baik lagi.

Bandung, Desember 2017

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi Penelitian	2
1.6 Sistematika Pembahasan	2
2 DASAR TEORI	5
2.1 Morfologi	5
2.2 Morfem	6
2.2.1 Identifikasi Morfem	6
2.2.2 Jenis Morfem	7
2.2.3 Morfem Dasar, Bentuk Dasar, Akar, dan Leksem	9
2.2.4 Morfem Afiks	10
2.3 Proses Morfologi	11
2.3.1 Bentuk Dasar	11
2.3.2 Pembentuk Kata	12
2.4 Morfofonemik	13
2.4.1 Jenis Perubahan	13
2.4.2 Prefiksasi ber-	13
2.4.3 Prefiksasi me- (termasuk klofiks me-kan dan me-i)	14
2.4.4 Prefiksasi pe- dan konfiksasi pe-an	15
2.4.5 Prefiksasi per- dan konfiksasi per-an	16
2.4.6 Prefiksasi ter-	16
2.5 Peran Morphological Parser dalam Natural Language Processing[1]	17
2.6 Struktur Data Trie	17
2.6.1 Bitwise Trie	18
2.6.2 Patricia Trie	18
2.6.3 Implementasi Trie dalam Kamus	19
2.6.4 Keunggulan Trie dibandingkan Struktur Data Lain	21
3 ANALISIS	23
3.1 Leksikon	23
3.2 Proses <i>Morphological Parsing</i>	25

3.3	Morfotaktik	26
3.4	Analisis Use Case	28
3.5	Analisis Kelas	31
4	PERANCANGAN	35
4.1	Struktur Penyimpanan Leksikon	35
4.2	<i>Syntax</i> Keluaran Proses Morphological Parsing	37
4.3	Perancangan Antarmuka	38
4.3.1	Perancangan Antarmuka Perangkat Lunak Morphological Parser	38
4.3.2	Perancangan Antarmuka Perangkat Lunak Lexicon	38
4.4	Perancangan Kelas Lengkap	43
4.4.1	Kelas Parser	44
4.4.2	Kelas Lexicon	48
4.4.3	Kelas Node	49
4.4.4	Kelas Combiner	50
4.5	Perancangan Diagram Aktivitas	50
4.6	Perancangan Algoritma	53
5	IMPLEMENTASI DAN PENGUJIAN	55
5.1	Implementasi	55
5.2	Pengujian	56
5.2.1	Pengujian Fungsional	56
5.2.2	Pengujian Nonfungsional	62
6	KESIMPULAN DAN SARAN	65
6.1	Kesimpulan	65
6.2	Saran	66
	DAFTAR REFERENSI	67
	A BAHAN PENGUJIAN	69
	B KODE PROGRAM	71

DAFTAR GAMBAR

2.1	Morfem bebas dan terikat	8
2.2	Morfem dasar dan afiks	9
2.3	Trie dengan kata "trie","ini", dan "contoh"	18
2.4	Pohon radix dengan 7 kata dengan prefix "r"	19
2.5	Trie dengan kata "hello","hey","number","name","cat","camel", dan "cup"	20
2.6	Penambahan kata "coding" pada trie di gambar 2.5	20
3.1	Struktur data trie	24
3.2	Diagram use case perangkat lunak morphological parser	29
3.3	Diagram kelas awal perangkat lunak morphological parser	33
4.1	Isi dari file sayur.lxc	36
4.2	Rancangan antarmuka perangkat lunak morphological parser	38
4.3	Rancangan antarmuka perangkat lunak lexicon halaman home untuk user	39
4.4	Rancangan antarmuka perangkat lunak lexicon halaman read untuk user	40
4.5	Rancangan antarmuka perangkat lunak lexicon halaman home untuk editor	40
4.6	Rancangan antarmuka perangkat lunak lexicon halaman create untuk editor	41
4.7	Rancangan antarmuka perangkat lunak lexicon halaman update untuk editor	42
4.8	Rancangan antarmuka perangkat lunak lexicon halaman delete untuk editor	43
4.9	Diagram kelas lengkap perangkat lunak morphological parser	44
4.10	Diagram aktivitas proses mengolah teks masukan	51
4.11	Diagram aktivitas proses parsing pada sebuah kata	52
5.1	Implementasi antarmuka perangkat lunak morphological parser	56
5.2	Implementasi antarmuka perangkat lunak lexicon	56
5.3	Hasil parsing contoh masukan pertama	57
5.4	Hasil parsing contoh masukan kedua	58
5.5	Hasil parsing contoh masukan pertama tanpa fitur converter	58
5.6	Hasil parsing contoh masukan kedua tanpa fitur validator	59
5.7	Isi leksikon sebelum kata 'warganet' dimasukkan	60
5.8	Isi leksikon setelah kata 'warganet' dimasukkan	60
5.9	Turunan dari kata dasar 'sayur' sebelum kata turunan 'menyayur-mayur' ditambahkan	61
5.10	Turunan dari kata dasar 'sayur' setelah kata turunan 'menyayur-mayur' ditambahkan	61
5.11	Isi leksikon sebelum kata 'abau' dihapus	62
5.12	Isi leksikon setelah kata 'abau' dihapus	62

DAFTAR TABEL

3.1	Tabel aturan morfotaktik untuk preprefiks	27
3.2	Tabel aturan morfotaktik untuk prefiks atau konfiks	28
3.3	Tabel aturan morfotaktik untuk akar	28
3.4	Tabel aturan morfotaktik untuk sufiks atau konfiks	28
4.1	Tabel lambang bentuk turunan dalam leksikon	36
5.1	Tabel waktu proses parsing untuk contoh masukan pertama	63
5.2	Tabel waktu proses parsing untuk contoh masukan kedua	63
5.3	Tabel waktu proses parsing untuk contoh masukan ketiga	63
5.4	Tabel waktu untuk melakukan create sebuah entri baru dalam leksikon	64
5.5	Tabel waktu untuk melakukan update sebuah entri dalam leksikon	64
5.6	Tabel waktu untuk melakukan delete sebuah entri dalam leksikon	64

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pengolahan bahasa alami atau dalam bahasa Inggris disebut dengan *natural language processing* (NLP) adalah cabang ilmu komputer dan linguistik yang mengkaji interaksi antara komputer dan manusia menggunakan bahasa alami. NLP sering dianggap sebagai cabang dari kecerdasan buatan dan bidang kajiannya bersinggungan dengan linguistik komputasional. Kajian NLP antara lain mencakup segmentasi tuturan (*speech segmentation*), segmentasi teks (*text segmentation*), penandaan kelas kata (*part-of-speech tagging*), serta pengawataksaan makna (*word sense disambiguation*). Salah satu alat yang digunakan oleh komputer dalam proses mengenali bahasa alami manusia adalah *morphological parser*.

Morphological parser berfungsi untuk membagi sebuah kata menjadi komponen-komponen penyusunnya. Proses ini dapat mengenali komponen kata seperti awalan, bentuk dasar, sisipan, dan akhiran serta dapat mengenali jika kata tersebut merupakan kata ulang maupun kata majemuk. Proses di mana *morphological parser* melakukan tugasnya dalam menguraikan kata menjadi komponen-komponen penyusunnya disebut dengan *morphological parsing*. Proses ini dapat membantu mengurangi ambiguitas selama proses mengetahui makna suatu kalimat. Sebagai contoh, kata "mengurus" bisa mempunyai makna menjadi kurus maupun mengerjakan sebuah urusan, bergantung pada apa bentuk dasar dari kata tersebut. Jika kita bisa membagi kata tersebut menjadi komponen penyusunnya, kita bisa lebih yakin mengenai makna dari kata tersebut dalam kalimat. *Morphological parsing* merupakan salah satu proses penting dalam NLP.

Morphological parser sudah banyak dibuat untuk beberapa bahasa yang ada di dunia, seperti bahasa Inggris, bahasa Turki, dan bahasa Bangla. Pisceldo et al. (2008) pernah membuat *morphological analyser* untuk bahasa Indonesia melalui pendekatan *two-level*, namun hanya dapat memproses kata hasil afiksasi dan reduplikasi. Dalam bahasa Indonesia, selain proses afiksasi dan reduplikasi, dikenal ada satu lagi proses morfologi yang umum dilakukan, yaitu proses komposisi. Proses komposisi adalah proses penggabungan bentuk dasar dengan bentuk dasar lain untuk mewadahi suatu "konsep" yang belum tertampung dalam sebuah kata[2]. Dalam skripsi ini, akan dibuat sebuah perangkat lunak *morphological parser* yang dapat memproses kata dalam bahasa Indonesia yang merupakan hasil proses afiksasi, reduplikasi, dan komposisi.

1.2 Rumusan Masalah

Sehubungan dengan latar belakang yang telah diuraikan di atas, maka dibuat rumusan masalah sebagai berikut ini.

- Bagaimana aturan morfologi bahasa Indonesia?
- Bagaimana struktur data dari *lexicon* yang digunakan pada perangkat lunak?
- Bagaimana cara mengimplementasikan aturan morfologi bahasa Indonesia ke dalam perangkat lunak?

- Bagaimana performansi dari perangkat lunak yang dihasilkan?

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

- Mengetahui aturan morfologi bahasa Indonesia
- Mengetahui struktur data dari *lexicon* yang digunakan pada perangkat lunak
- Mengimplementasikan aturan morfologi bahasa Indonesia ke dalam perangkat lunak
- Mengetahui performansi dari perangkat lunak yang dihasilkan

1.4 Batasan Masalah

Terdapat beberapa batasan masalah untuk penelitian ini:

- Kalimat yang dapat diproses adalah kalimat dalam bahasa Indonesia yang ditulis sesuai ejaan yang disempurnakan (EYD)
- Kata yang dapat diproses adalah kata yang merupakan bentuk dasar dan kata yang dibentuk dari proses morfologi berupa afiksasi, reduplikasi, dan komposisi
- Kata yang belum ada dalam Kamus Besar Bahasa Indonesia (KBBI) dan yang bukan merupakan hasil dari proses afiksasi, reduplikasi, dan komposisi dianggap sebagai bentuk asing
- Kata yang merupakan hasil proses penyisipan (infiksasi) dan belum ada dalam KBBI tidak dapat diproses karena infiksasi dianggap sudah tidak produktif dalam bahasa Indonesia pada saat ini

1.5 Metodologi Penelitian

Tahap-tahap yang akan dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan studi literatur tentang morfologi bahasa Indonesia dan perangkat lunak *morphological parser* yang sudah ada
2. Melakukan analisis pada *morphological parser* bahasa Indonesia dan *lexicon* yang digunakan serta merancang struktur data dari *lexicon*
3. Merancang dan mengimplementasikan *lexicon* dan *morphological parser* ke dalam perangkat lunak
4. Mengumpulkan contoh kalimat dalam bahasa Indonesia sebagai bahan pengujian
5. Melakukan pengujian terhadap perangkat lunak

1.6 Sistematika Pembahasan

Keseluruhan bab yang disusun dalam karya tulis ini terbagi ke dalam bab-bab sebagai berikut:

1. BAB 1 - PENDAHULUAN membahas mengenai latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.

-
2. BAB 2 - DASAR TEORI membahas mengenai teori-teori dasar yang digunakan pada penelitian ini.
 3. BAB 3 - ANALISIS membahas mengenai hasil analisis dari teori dasar dan kebutuhan dari perangkat lunak yang akan dibuat.
 4. BAB 4 - PERANCANGAN membahas mengenai perancangan perangkat lunak berdasarkan hasil analisis yang telah dilakukan.
 5. BAB 5 - IMPLEMENTASI DAN PENGUJIAN membahas mengenai implementasi dari hasil analisis dan perancangan perangkat lunak dan pengujian terhadap perangkat lunak yang sudah dibuat.
 6. BAB 6 - KESIMPULAN DAN SARAN membahas mengenai kesimpulan dan saran dari hasil penelitian.