

**POS TAGGING BAHASA INDONESIA BERBASIS ATURAN
DAN LEXICON (KATA KERJA + KATA KETERANGAN)**

Skripsi



oleh

YEDIJA KADMIEL ELNATAN

71130024

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2017

**POS TAGGING BAHASA INDONESIA BERBASIS ATURAN
DAN LEXICON (KATA KERJA + KATA KETERANGAN)**

Skripsi



©
Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

YEDIJA KADMIEL ELNATAN

71130024

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2017

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

POS TAGGING BAHASA INDONESIA BERBASIS ATURAN DAN LEXICON (KATA KERJA + KATA KETERANGAN)

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 17 Oktober 2017



YEDIJA KADMIEL ELNATAN
71130024

HALAMAN PERSETUJUAN

Judul Skripsi : POS TAGGING BAHASA INDONESIA BERBASIS
ATURAN DAN LEXICON (KATA KERJA + KATA
KETERANGAN)

Nama Mahasiswa : YEDIJA KADMIEL ELNATAN

N I M : 71130024

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2017/2018

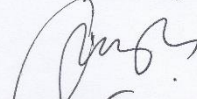
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 17 Oktober 2017

Dosen Pembimbing I



Lucia Dwi Krisnawati, Dr.

Dosen Pembimbing II



Kathryn Widhiyanti, M.Cs.

HALAMAN PENGESAHAN

POS TAGGING BAHASA INDONESIA BERBASIS ATURAN DAN LEXICON (KATA KERJA + KATA KETERANGAN)

Oleh: YEDIJA KADMIEL ELNATAN / 71130024

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 17 Oktober 2017

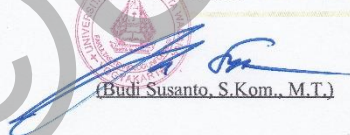
Yogyakarta, 17 Oktober 2017
Mengesahkan,

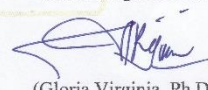
Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr.
2. Kathryn Widhiyanti, M.Cs.
3. Yuan Lukito, S.Kom., M.Cs.
4. Danny Sebastian, S.Kom., M.M., M.T.

Dekan

Ketua Program Studi


(Budi Susanto, S.Kom., M.T.)


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Dalam penelitian tugas akhir ini, penulis mendapatkan bantuan, saran, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Budi Susanto, S.Kom., M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia, S.Kom., MAI., Ph.D. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
3. Ibu Dr. Lucia Dwi Krisnawati, selaku dosen pembimbing I yang telah memberikan waktu secara rutin untuk melakukan konsultasi dan memberikan saran dan masukan mengenai pemrograman sistem, dan penyelesaian masalah dengan cara yang lebih sederhana.
4. Ibu Kathryn Widhiyanti, M.Cs. selaku dosen pembimbing II yang telah meluangkan waktu untuk memberikan bimbingan dan memberikan saran serta masukan mengenai penulisan laporan juga analisis sistem.
5. Stefanus Candra Saputra, S.Pd., Herningdyah Cahyaning Ratri, S.Pd., dan Lorensia Dwi Mawar yang telah membantu dalam pengerjaan dokumen evaluasi
6. Orangtua, saudara-saudara dan teman-teman terdekat yang selalu memberikan dukungan motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
7. Pihak – pihak lain yang tidak dapat penulis sebutkan satu per satu yang berperan secara langsung maupun tidak langsung selama pengerjaan tugas akhir.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan kasih karunia-Nya sehingga penulis dapat menyelesaikan pembuatan sistem dan laporan tugas akhir dengan judul “*POS TAGGING* BAHASA INDONESIA BERBASIS ATURAN DAN *LEXICON* (KATA KERJA + KATA KETERANGAN)” dengan baik.

Penulisan laporan tugas akhir ini diajukan sebagai salah satu syarat guna mencapai gelar Sarjana Strata Satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta.

Dalam pembuatan laporan ini, penulis menyadari bahwa masih ada kekurangan, baik dari materi maupun teknik penyajiannya. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran dari pembaca. Akhir kata penulis memohon maaf apabila dalam penulisan laporan ini, ada kalimat yang kurang berkenan. Semoga hasil dari pengerjaan tugas akhir ini dapat berguna dan bermanfaat bagi banyak pihak.

Yogyakarta, 17 Oktober 2017

Penulis

INTISARI

POS TAGGING BAHASA INDONESIA BERBASIS ATURAN DAN *LEXICON* (KATA KERJA + KATA KETERANGAN)

Part-of-Speech Tagging (POS Tagging) merupakan proses pemberian label kelas kata secara otomatis pada suatu kata dalam kalimat. Tujuan dari perancangan sistem ini adalah untuk membantu proses *POS Tagging* terhadap dokumen Bahasa Indonesia. Pemberian label kata dapat dilakukan dengan berbagai metode, yaitu metode berbasis aturan, metode statistik, dan metode *transformation based learning*.

Pada penelitian ini, penulis menggunakan metode berbasis aturan dan leksikon untuk membedakan kelas kata kerja dan kelas kata keterangan. *Preprocessing* dalam penelitian ini menggunakan normalisasi, tokenisasi, dan *case folding*. Proses pengecekan dilakukan dengan menggunakan aturan-aturan, seperti aturan duplikasi, aturan imbuhan awalan, aturan imbuhan akhiran, aturan khusus, dan pengecekan kata dasar. Hasil *output* sistem berupa kata yang telah diberikan *tag* sesuai dengan kelas katanya.

Hasil menunjukkan rata-rata akurasi sebesar 99.3%. Setelah memperbaiki dan menambah leksikon, akurasi meningkat menjadi 99.6% yang mengindikasikan metode berbasis aturan baik mengenali kelas kata kerja dan kelas kata keterangan dalam teks Bahasa Indonesia. Kelemahan sistem adalah sistem belum bisa mendeteksi kata verba majemuk, duplikasi fonologi, kata ambiguitas, dan sistem juga tidak melihat pola kata dalam *input*, sehingga hasil yang terbentuk dari kata kerja dan kata keterangan kadang tidak sesuai dengan konteks makna dalam kalimat.

Kata Kunci: [*POS Tagging*, Bahasa Indonesia, berbasis aturan, leksikon, *tag*]

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR	vii
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xv
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Metodologi Penelitian	4
1.6. Sistematika Penulisan.....	5
BAB 2 TINJAUAN PUSTAKA.....	7
2.1. Tinjauan Pustaka	7
2.2. Landasan Teori	9
2.2.1. Kelas Kata	9
2.2.2. Kata Verba	10
2.2.3. Kata Adverbia	12
2.2.4. Imbuhan / Afiksasi	13
2.2.5. <i>Text Preprocessing</i>	18
2.2.6. <i>POS Tagging</i>	20
2.2.7. <i>Rule Based Part-Of-Speech Tagging</i>	20
2.2.8. <i>Lexicon</i>	21
2.2.9. <i>Tagset</i>	21
2.2.10. Evaluasi	23

BAB 3 PERANCANGAN SISTEM.....	24
3.1. Tahapan Pembangunan Sistem.....	24
3.1.1. Pra-pemrosesan	24
3.1.2. Pemrosesan.....	25
3.1.3. Pasca Pemrosesan.....	25
3.2. Analisis Kebutuhan Sistem	25
3.2.1. Kebutuhan Fungsional	26
3.2.2. Kebutuhan Non – Fungsional.....	26
3.2.3. Kebutuhan Hardware dan Software	26
3.3. Rancangan Sistem	27
3.3.1. Alur Kerja Sistem.....	27
3.3.2. Alur Kerja Sistem <i>Preprocessing</i>	28
3.3.3. Alur Kerja Sistem Berbasis Aturan.....	28
3.3.4. <i>Use Case Diagram</i>	29
3.3.5. Antarmuka.....	32
3.3.6. Basis Data	38
3.4. Kamus Data	39
3.5. Rancangan Pengujian	39
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	41
4.1. Implementasi Sistem	41
4.1.1. Implementasi Pembuatan Leksikon	41
4.1.2. Implementasi Aturan Kata Kerja dan Kata Keterangan.....	42
4.1.3. Implementasi Aturan Khusus.....	43
4.1.4. Implementasi Alur Kerja Sistem Berbasis Aturan	44
4.1.5. Implementasi Pembuatan Dokumen Evaluasi.....	50
4.1.6. Implementasi Basis Data.....	50
4.1.7. Implementasi Antarmuka	52
4.2. Implementasi Program	63
4.2.1. Proses <i>Preprocessing</i>	63
4.2.2. Proses Pengecekan Aturan dan Pemberian <i>Tag</i>	65
4.3. Hasil Pengujian Sistem.....	69
4.3.1. Presisi	69
4.3.2. <i>Recall</i>	72
4.3.3. Akurasi	76

4.4. Analisis Pengujian Sistem	78
4.4.1. Pembahasan Analisis.....	78
4.4.2. Kekurangan Sistem	83
BAB 5 KESIMPULAN DAN SARAN	84
5.1. Kesimpulan.....	84
5.2. Saran.....	84
DAFTAR PUSTAKA	86
LAMPIRAN A	1
LAMPIRAN B	1
LAMPIRAN C	1

©UKDWN

DAFTAR TABEL

Tabel 2.1. Tagset II Bahasa Indonesia	21
Tabel 3.1. Use Case Input data.....	30
Tabel 3.2. Use Case Simpan data.....	31
Tabel 3.3. Use Case CRUD Leksikon.....	31
Tabel 3.4. Use Case CRUD Leksikon.....	31
Tabel 3.5. Use Case CRUD Leksikon.....	32
Tabel 3.6. Use Case Evaluasi	32
Tabel 3.7. Kamus Data (Data Dictionary)	39
Tabel 3.8. Confusion Matrix	40
Tabel 4.1. Pengujian Dokumen Uji 1 dan Dokumen Evaluasi 1	69
Tabel 4.2. Hasil Presisi Pengujian Sistem.....	70
Tabel 4.3. Pengujian Dokumen Uji 1 dan Dokumen Evaluasi 1	72
Tabel 4.4. Hasil Recall Pengujian Sistem	74
Tabel 4.5. Pengujian Dokumen Uji 1 dan Dokumen Evaluasi 1	76
Tabel 4.6. Hasil Akurasi Pengujian Sistem.....	76
Tabel 4.7. Hasil Pengujian Sistem	78
Tabel 4.8. Daftar Kesalahan Pemberian Tag pada Kata	79
Tabel 4.9. Hasil Pengujian Ulang Presisi, Recall, Akurasi	81
Tabel 4.10. Hasil Pengujian Ulang Sistem.....	82

DAFTAR GAMBAR

Gambar 2.1. Penggolongan Kata Verba Berdasarkan Bentuk	11
Gambar 2.2. Penggolongan Kata Adverbia Berdasarkan Bentuk	12
Gambar 2.3. Text Preprocessing	19
Gambar 2.4. Proses Tokenisasi	19
Gambar 2.5. Contoh Rule Based Part-Of-Speech Tagging.....	20
Gambar 2.6. Daftar kata yang tersimpan dalam lexicon	21
Gambar 2.7. Confusion Matrix	23
Gambar 3.1. Alur Kerja Sistem.....	28
Gambar 3.2. Tahap Preprocessing	28
Gambar 3.3. Proses Aturan	29
Gambar 3.4. Use Case Diagram sistem POS Tagging Bahasa Indonesia.....	30
Gambar 3.5. Antarmuka Halaman Beranda	33
Gambar 3.6. Antarmuka Halaman POS	33
Gambar 3.7. Antarmuka Halaman POS Output	34
Gambar 3.8. Antarmuka Halaman Tentang Kami	34
Gambar 3.9. Antarmuka Halaman Bantuan	35
Gambar 3.10. Antarmuka Halaman Login	35
Gambar 3.11. Antarmuka Halaman Leksikon Bukan Kata Kerja.....	36
Gambar 3.12. Antarmuka Halaman Leksikon Kata Dasar.....	36
Gambar 3.13. Antarmuka Halaman Evaluasi.....	37
Gambar 3.14. Antarmuka Halaman Evaluasi Output.....	37
Gambar 3.15. Antarmuka Halaman Admin	38
Gambar 3.16. Rancangan Basis Data	38
Gambar 4.1. Implementasi Alur Kerja Sistem Berbasis Aturan	44
Gambar 4.2. Implementasi Alur Kerja Tahap Reduplikasi.....	45
Gambar 4.3. Implementasi Alur Kerja Imbuhan Awalan Kata Kerja.....	47
Gambar 4.4. Implementasi Alur Kerja Imbuhan Akhiran Kata Kerja	48
Gambar 4.5. Implementasi Alur Kerja Awalan dan Akhiran Kata Keterangan....	49
Gambar 4.6. Implementasi Alur Kerja Kata Dasar	49
Gambar 4.7. Desain Basis Data	51
Gambar 4.8. Contoh isi tabel bukan_kata_kerja	51
Gambar 4.9. Contoh isi tabel kata_dasar	52
Gambar 4.10. Contoh isi tabel user	52
Gambar 4.11. Halaman Awal Sistem	53
Gambar 4.12. Tampilan menu POS	54
Gambar 4.13. Tampilan menu POS Output	54
Gambar 4.14. Tampilan menu Tentang Kami.....	55
Gambar 4.15. Tampilan menu Bantuan	55
Gambar 4.16. Halaman menu Login	56
Gambar 4.17. Tampilan menu Admin.....	57

Gambar 4.18. Tampilan menu Leksikon Kata Dasar	57
Gambar 4.19. Tampilan menu Leksikon Bukan Kata Kerja	58
Gambar 4.20. Tampilan menu Evaluasi	58
Gambar 4.21. Tampilan menu Evaluasi Output	59
Gambar 4.22. Tampilan Tambah Leksikon Kata Dasar	59
Gambar 4.23. Tampilan Tambah Leksikon Bukan Kata Kerja	60
Gambar 4.24. Tampilan Tambah Admin	60
Gambar 4.25. Tampilan Edit Leksikon Kata Dasar	61
Gambar 4.26. Tampilan Edit Leksikon Bukan Kata Kerja	61
Gambar 4.27. Tampilan Edit Admin	61
Gambar 4.28. Tampilan Hapus Leksikon Kata Dasar	62
Gambar 4.29. Tampilan Hapus Leksikon Bukan Kata Kerja	62
Gambar 4.30. Tampilan Hapus Admin	62
Gambar 4.31. Tampilan Menu Pencarian	63
Gambar 4.32. Grafik Perhitungan Macro-Average Presisi	72
Gambar 4.33. Grafik Perhitungan Macro-Average Recall	75

©UKDW

DAFTAR LAMPIRAN

Lampiran A	A-1
Lampiran B.....	B-1
Lampiran C.....	C-1

©UKDW

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Dengan perkembangan teknologi yang semakin pesat, setiap orang dituntut untuk bisa memanfaatkan dengan baik perkembangan teknologi dan dapat menggunakan di dalam kehidupan sehari-hari. Perkembangan teknologi mencakup semua aspek kehidupan, salah satunya adalah bahasa. Bahasa merupakan alat komunikasi untuk antar manusia agar dapat berkomunikasi dan berinteraksi satu dengan yang lainnya. Itulah mengapa bahasa menjadi salah satu faktor krusial dalam kehidupan bermasyarakat di dunia. Di era teknologi yang semakin maju ini, *machine learning* membutuhkan pengetahuan mengenai bahasa, khususnya dalam hal pembagian kelas kata. Yang menjadi permasalahan selama ini adalah bagaimana mesin dapat mengerti bahasa yang kita inginkan, sehingga mesin dapat memproses sesuai dengan keinginan pengguna. Mesin dapat mengerti bahasa yang kita inginkan karena adanya pengelompokan kelas kata. Contohnya adalah mesin penerjemah kalimat. Mesin dapat mengelompokkan secara otomatis kata “*bisa*” untuk kata benda ataupun kata kerja. Cara mengajarkan bahasa ke mesin dapat dilakukan dengan berbagai metode. Metode itu meliputi *rule based*, statistik, dan *transformation based learning*. Dengan menggunakan metode-metode tersebut, nantinya dapat mencapai pelabelan kata. Dan salah satu pelabelan kata adalah *POS Tagging*.

Salah satu pemanfaatan teknologi dalam bidang bahasa adalah adanya *POS Tagging*. *POS Tagging (Part-Of-Speech Tagging)* adalah proses penandaan untuk menempatkan *part-of-speech* atau penanda kelas leksikal lain untuk setiap kata dalam korpus (Jurafsky & Martin, 2000). Contoh penggunaan *POS Tagging*, “*saya makan nasi*” menjadi *saya/KG makan/V nasi/NN*, dimana label **KG** = kata ganti, **V** = kata kerja, **NN** = kata benda (Christanti, Pragantha, & Purnamasari, 2012).

Hasil penelitian dari *Part-of-Speech Tagging* pada dokumen dapat digunakan sebagai dasar penelitian dalam *Natural Language Processing* lainnya, seperti: *Language Generator*, *Information Retrieval*, *Text Summarization*, *Question and Answering*, dan *Machine Translation*. *POS Tagging (Part-Of-Speech Tagging)* sangat membantu dalam pengelompokan jenis – jenis kata.

Dalam hal ini, penulis ingin membuat suatu penelitian mengenai *POS Tagging (Part-Of-Speech Tagging)* khususnya Bahasa Indonesia. Dalam penelitian ini, metode yang akan digunakan adalah metode berbasis aturan. Hal ini dilakukan karena belum adanya korpus yang dilabeli dengan kelas kata, dan tidak tersedianya korpus yang diberikan secara umum. Dalam penelitian ini, kata yang akan dianalisa adalah kata kerja dan kata keterangan saja. Hal ini dilakukan karena, menurut penulis kata kerja dan kata keterangan merupakan kata yang saling mendukung dalam suatu kalimat, contohnya, “ibu baru saja pulang dari pasar”. Dalam kalimat tersebut, kata “pulang” termasuk dalam kata kerja, sedangkan kata “saja” merupakan kata keterangan. Kata “baru” sebenarnya masuk kelas adjektiva, namun frasa “baru saja” masuk dalam kategori frasa keterangan. Dan frasa tersebut saling mendukung untuk membentuk satu kalimat. Pemilihan menggunakan kelas kata kerja dan kelas kata keterangan, dikarenakan ruang lingkup untuk kelas kata cukup banyak, aturan yang akan dihasilkan sangat banyak. Sehingga kelas kata yang dipilih untuk penelitian ini adalah kelas kata kerja dan kata keterangan.

1.2. Rumusan Masalah

Berdasarkan masalah yang dijabarkan pada latar belakang, maka perumusan masalah yang ada antara lain yaitu:

- a) Bagaimana menyusun aturan – aturan *POS Tagging (Part-Of-Speech Tagging)* kelas kata bahasa Indonesia untuk kelas kata kerja dan kelas kata keterangan?
- b) Bagaimana sistem dapat memberikan label kelas kata yaitu kelas kata kerja dan kelas kata keterangan dalam sebuah kalimat?

- c) Berapa persen tingkat akurasi sistem mengenali kata dengan label kata kerja dan kata keterangan?

1.3. Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yang digunakan, yaitu:

- a) *Input* pada sistem ini hanya berbentuk teks atau *file* yang berekstensi *.txt.
- b) Bahasa yang digunakan bahasa Indonesia.
- c) Metode yang akan digunakan menggunakan adalah metode berbasis aturan.
- d) Sistem hanya akan memproses kata kerja dan kata keterangan.
- e) Dokumen uji dan dokumen evaluasi yang dipakai total berjumlah 80 dokumen. Masing-masing 40 dokumen untuk dokumen uji dan 40 dokumen untuk dokumen evaluasi.
- f) Dokumen uji yang akan dipakai adalah artikel – artikel berita yang menggunakan bahasa Indonesia baku.
- g) Dokumen evaluasi yang akan digunakan telah *tag* secara manual oleh orang yang ahli di bidang bahasa Indonesia.
- h) *Output* yang akan dihasilkan dari sistem ini berupa kata yang telah diberi label kata kerja dan kata keterangan dalam suatu kalimat.
- i) Leksikon yang digunakan adalah leksikon bukan kata kerja dan leksikon kata dasar.
- j) Sistem yang akan dibangun menggunakan bahasa pemrograman PHP yang diterapkan dalam bentuk web.

1.4. Tujuan Penelitian

Melalui penelitian ini, tujuan yang ingin dicapai adalah:

- a) Menghasilkan sistem *Tagger* berbasis aturan bagi kelas kata kerja dan kelas kata keterangan.

- b) Membangun korpus bahasa Indonesia yang terdiri dari kata yang telah dilabeli dengan kata kerja dan kata keterangan.
- c) Mengetahui tingkat akurasi keluaran sistem *POS Tagging (Part-Of-Speech Tagging)* berbasis aturan untuk dokumen bahasa Indonesia.

1.5. Metodologi Penelitian

Proses penelitian *POS Tagging (Part-Of-Speech Tagging)* Bahasa Indonesia memiliki beberapa langkah. Langkah – langkah yang akan dilakukan adalah sebagai berikut:

a) Tahap Pertama (Studi Pustaka)

Studi pustaka dilakukan dengan cara mempelajari teori - teori yang berkaitan dengan *POS Tagging*, metode berbasis aturan, *lexicon*, kelas kata dalam bahasa Indonesia melalui beberapa buku, jurnal, artikel, dan bahan lain yang mendukung untuk penelitian ini.

b) Tahap Kedua (Pengumpulan data dan pembuatan leksikon)

Pengumpulan data diambil dari artikel – artikel berita *online* yang menggunakan bahasa Indonesia yang baku. Jenis artikel yang akan digunakan bermacam - macam, yaitu politik, olahraga, ekonomi, dan masih banyak lagi. Sumber yang akan digunakan untuk artikel *online*, seperti <http://www.kompas.com/>, <https://www.detik.com/>. Sedangkan untuk pembuatan leksikon, mengambil dari Kamus Besar Bahasa Indonesia (KBBI) serta dari internet.

c) Tahap Ketiga (Implementasi atau *coding*)

Tahap ini merupakan tahap pengerjaan sistem untuk mendapatkan kelas kata kerja dan kelas kata keterangan. Sistem akan menghasilkan *output* berupa kata yang telah terlabeli dengan kelas kata kerja dan kelas kata keterangan.

d) Tahap Keempat (Evaluasi sistem)

Tahap evaluasi akan dilakukan dengan cara membandingkan hasil dokumen yang telah diproses oleh sistem dengan hasil dokumen yang telah dievaluasi oleh

orang yang membidangi ilmu bahasa. Dari tahap evaluasi, nantinya akan didapatkan nilai presisi, *recall*, dan akurasi.

e) Tahap Kelima (Kesimpulan)

Pada tahap ini akan ditarik kesimpulan berdasarkan hasil dari tingkat akurasi dari hasil evaluasi yang telah dilakukan.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan penulis untuk mengerjakan laporan tugas akhir ini sesuai dengan ketentuan dari buku panduan penulisan tugas akhir. Penulisan akan dibagi menjadi 5 bab, yaitu :

Bab I : Pendahuluan

Pada bab pertama akan dibahas mengenai latar belakang, rumusan masalah, batasan dari sistem yang akan dibangun, tujuan yang ingin dicapai oleh penulis, dan metodologi penelitian.

Bab II : Tinjauan Pustaka

Pada bab kedua akan menjelaskan mengenai tinjauan pustaka dan landasan teori. Pada tinjauan pustaka akan berisi penelitian-penelitian yang sudah pernah dilakukan oleh orang lain berkaitan dengan penelitian yang akan dilakukan oleh penulis. Sedangkan landasan teori akan berisi tentang dasar-dasar teori yang akan digunakan pada penelitian ini, termasuk metode-metode yang akan digunakan.

Bab III : Perancangan Sistem

Pada bab ketiga penulis membahas mengenai perancangan sistem yang akan dibuat. Bab ini menjelaskan bagaimana tahapan-tahapan pembangunan sistem, analisis kebutuhan sistem, rancangan sistem, serta rancangan pengujian dari sistem.

Bab IV : Implementasi dan Analisis Sistem

Pada bab keempat penulis akan membahas mengenai hasil implementasi yang telah dilakukan. Hasil yang akan didapatkan dapat berupa hasil riset atau penelitian yang telah dilakukan oleh penulis.

Bab V : Kesimpulan

Pada bab kelima penulis akan membahas mengenai kesimpulan dari hasil penelitian dan saran. Kesimpulan berisi tentang hasil yang telah didapatkan selama penelitian berlangsung. Sedangkan saran dapat berupa sesuatu hal yang dapat dilakukan berkaitan dengan pengembangan topik penelitian tersebut di masa mendatang serta metode yang dapat digunakan.

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil analisis penelitian yang telah dibahas pada Bab 4, maka dapat ditarik kesimpulan sebagai berikut:

1. Sistem ini sudah cukup baik dalam mengenali kelas kata kerja dan kelas kata keterangan. Pada pengujian pertama rata-rata akurasi sistem mencapai 99,3%. Setelah melakukan evaluasi dan melakukan perbaikan dengan cara memperbaiki leksikon kata dasar dan memperkaya daftar leksikon bukan kata kerja, rata-rata akurasi meningkat menjadi 99.6%. Hal ini menunjukkan bahwa metode berbasis aturan baik dalam mengenali kelas kata, khususnya kelas kata kerja dan kelas kata keterangan.
2. Penyusunan aturan dimulai dari aturan reduplikasi, aturan berimbuhan awalan, aturan berimbuhan akhiran, aturan khusus, dan pengecekan kata dasar. Setiap kata akan melalui proses aturan dan hasilnya berupa kata beserta *tag* yang tepat dan sesuai.
3. Kelemahan dari sistem ini adalah sistem belum bisa menyelesaikan kata yang termasuk verba majemuk, reduplikasi fonologi, kata ambiguitas. Sistem juga tidak melihat pola kata dalam *input*, sehingga hasil yang terbentuk dari kata kerja dan kata keterangan kadang tidak sesuai dengan konteks makna dalam kalimat.

5.2. Saran

Sistem ini sangat memungkinkan untuk dilakukan pengembangan lebih lanjut sesuai kebutuhan yang terus bertambah, sehingga dapat meningkatkan akurasi

sistem. Saran yang diajukan penulis dalam pengembangan sistem kedepannya adalah sebagai berikut:

1. Memperbaiki dan memperkaya daftar kata-kata yang terdapat di dalam leksikon yang digunakan. Sehingga kata mendapatkan *tag* yang tepat dan sesuai dengan kelas katanya.
2. Pada penelitian ini, *POS Tagging* dilakukan hanya untuk mengenali kelas kata kerja dan kelas kata keterangan. Penulis mengharapkan agar program terus dikembangkan pada setiap bentuk kelas kata dalam bahasa Indonesia. Sehingga nantinya dapat dikembangkan menjadi aplikasi mesin penerjemah yang tepat dan akurat.
3. Melakukan analisa sintaksis dan pemberian pembobotan berdasarkan frekuensi dalam sebuah korpus, agar sistem mampu mengatasi masalah ambiguitas, kata majemuk, dan reduplikasi fonologi. Sehingga sistem yang dihasilkan mampu memberikan *output* kelas kata yang benar-benar tepat dan sesuai dengan konteks makna dalam kalimat.

DAFTAR PUSTAKA

- Adikara, P. P. (2012, November 17). *Kamus Kata Dasar dan Stopword List Bahasa Indonesia*. Retrieved from Putra Pandu Adikara のブログ: <http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/>
- Alwi, H., & dkk. (2003). *Tata Bahasa Baku Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Chaer, A. (2006). *Tata Bahasa Praktis Bahasa Indonesia*. Jakarta: PT Rineka Cipta.
- Chaer, A. (2008). *Morfologi Bahasa Indonesia (Pendekatan Proses)*. Jakarta: PT Rineka Cipta.
- Christanti, V. M., Pragantha, J., & Purnamasari, E. (2012). IMPLEMENTASI BRILL TAGGER UNTUK MEMBERIKAN POS-TAGGING PADA DOKUMEN BAHASA INDONESIA. *Jurnal Teknik dan Ilmu Komputer*, 1(3), 301-315.
- DATAQ. (2013, Juni 16). *Perbedaan: precision, recall & accuracy*. Retrieved from DATA's BASE: <https://dataq.wordpress.com/2013/06/16/perbedaan-precision-recall-accuracy/>
- Departemen Pendidikan Nasional. (2008). *Kamus Besar Bahasa Indonesia Pusat Bahasa Edisi Keempat*. Jakarta: PT Gramedia.
- Gurusamy, V., & Kannan, S. (2014). Preprocessing Techniques for Text Mining. *ResearchGate*, 1-7.
- Jurafsky, D., & Martin, J. (2000). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics*. New Jersey: Alan Apt.

- Karibun, M. (2005). PENERAPAN METODE RULE-BASED DENGAN UNSUPERVISED LEARNING UNTUK PELABELAN DOKUMEN BERBAHASA LEARNING UNTUK PELABELAN DOKUMEN BERBAHASA.
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2016). *KBBI Daring*. Retrieved from KBBI Daring: <https://kbbi.kemdikbud.go.id/>
- Kridalaksana, H. (2005). *Kelas Kata dalam Bahasa Indonesia*. Jakarta: PT Gramedia Pustaka Utama.
- Pisceldo, F., Adriani, M., & Manurung, R. (2009). Probabilistic Part Of Speech Tagging for Bahasa Indonesia.
- Sabloak, N., Hardono, B. A., & Alamsyah, D. (2016). Part-of-Speech (POS) Tagging Bahasa Indonesia Menggunakan Algoritma Viterbi.
- Setiawan, E. (2017). *Kamus Besar Bahasa Indonesia (KBBI)*. Retrieved from Kamus Besar Bahasa Indonesia (KBBI): <https://kbbi.web.id/>
- Sundari, P., Zuraiyah, T. A., & Maryana, S. (n.d.). PENERAPAN RULE BASED SYSTEM PADA KAMUS BAHASA INDONESIA-MINANGKABAU BERBASIS ANDROID.
- Tim Prima Pena. (n.d.). *Kamus Besar Bahasa Indonesia Edisi Terbaru*. Gitamedia Press.
- Widhiyanti, K., & Harjoko, A. (2012). POS Tagging Bahasa Indonesia Dengan HMM dan Rule Based. *INFORMATIKA*, 8(2), 151-167.