

## PENENTUAN KELAS KATA PADA PART OF SPEECH TAGGING KATA AMBIGU BAHASA INDONESIA

Ahmad Subhan Yazid<sup>(1)</sup>, Agung Fatwanto<sup>(2)</sup>

Teknik Informatika Fakultas Sains dan Teknologi, UIN Sunan Kalijaga Yogyakarta

Jl. Laksda Adisucipto, Yogyakarta 55281

e-mail : subhan.yazid@uin-suka.ac.id<sup>(1)</sup> , agung.fatwanto@uin-suka.ac.id<sup>(2)</sup>,

### ABSTRACT

*Indonesian is in ninth as the language with the largest number of speakers in the world. hold a fundamental role in the communication. However, there is a problem in its application to the learning machine, which is about ambiguity (multi interpretation). In the Natural Language Processing study, Part of Speech (POS) tagging stage has a role in the decreasing this problem. This study use the Rule Base method to determine the best word class for ambiguous words in Indonesian. The research follows some stages: knowledge inventory, making algorithms, implementation, testing, and analysis. The first data used is Indonesian corpus that was developed by Language department of Computer science Faculty, Indonesia University. Then, data is processed and shown descriptively by following certain rules and specification. The result is a POS tagging algorithm included 71 rules in flowchart and descriptive sentence notation. Refer to testing result, the algorithm successfully provides 92 labeling of 100 tested ambiguous words (92%). The results of the implementation are influenced by the availability of rules, word class tagsets and corpus data.*

**Keywords:** Algorithm, Ambiguous, POS tagging, Rule based.

### ABSTRAK

Bahasa Indonesia menempati peringkat kesembilan sebagai bahasa dengan jumlah penutur terbanyak di dunia. Namun demikian, terdapat permasalahan dalam penerapannya pada mesin pembelajaran, yakni tentang ambiguitas (multi interpretasi). Dalam kajian Pemrosesan Bahasa Alami, tahapan Part-of-Speech (POS) tagging berperan dalam mengurangi permasalahan ini. Penelitian ini menggunakan metode Rule Base untuk menentukan kelas kata pada POS tagging kata ambigu bahasa Indonesia. Metode penelitian yang ditempuh adalah: inventarisasi pengetahuan, penyusunan algoritma, implementasi, pengujian, dan analisis. dan penarikan kesimpulan. Data awal yang digunakan pada penelitian ini adalah korpus Bahasa Indonesia yang dikembangkan oleh Divisi Bahasa fakultas ilmu komputer Universitas Indonesia. Data kemudian diolah dan disajikan secara deskriptif dengan mengacu pada beberapa batasan dan spesifikasi tertentu. Hasilnya, didapatkan sebuah algoritma POS tagging dengan 71 aturan yang dinotasikan dalam kalimat deskriptif dan diagram alir (flowchart). Berdasarkan dari hasil pengujian, algoritma tersebut dapat menangani 92 dari 100 kata ambigu (92%). Beberapa faktor yang mempengaruhi kinerja algoritma ini adalah kelengkapan aturan, label kelas kata dan data korpus.

**Kata Kunci:** Algoritma, Ambigu, POS tagging, Rule Based

### 1. PENDAHULUAN

Bahasa Indonesia, termasuk bahasa lainnya, berfungsi sebagai alat komunikasi antar manusia dalam menyalurkan informasi, perasaan, sikap, gagasan, emosi atau tekanan satu sama lain. Bahasa Indonesia menjadi hal yang substansial, sebab ia merupakan jembatan yang mempersatukan berbagai bahasa yang ada di Indonesia sebagai relasi atas tingginya keragaman budaya dan suku yang ada di dalamnya. Data statistik menyebutkan, pengguna bahasa Indonesia adalah 260 juta orang di seluruh dunia (Na'im Syaputra, 2011). Jumlah ini menempatkan bahasa Indonesia sebagai bahasa dengan jumlah penutur terbanyak kesembilan di dunia (Solehudin, 2014). Betapapun demikian, ada permasalahan yang muncul dalam penggunaan bahasa In-

donesia, yakni tentang ambiguitas. Ambiguitas adalah sifat atau hal yang memiliki pengertian atau bermakna lebih dari satu (Tim-Bahasa, 2008). Contoh sederhana ambiguitas adalah penggunaan kata bisa pada kalimat Joni bisa terkena bisa ular. Penggunaan kata bisa yang pertama merujuk pada kata kerja yang berarti mampu atau dapat, sedangkan bisa yang kedua merujuk pada zat racun. Dalam bahasa lisan, penafsiran kalimat ambigu ini mungkin tidak akan terjadi karena adanya perbedaan dalam intonasi pengucapannya. Namun, dalam bahasa tulis (dituliskan), penafsiran ganda ini dapat saja terjadi jika penanda-penanda ejaan tidak dilengkapi dengan petunjuk. Lebih jauh, ambiguitas ini dapat menimbulkan ketidakakuratan dalam penentuan label kata yang dapat berdampak pada kesalahan dalam memaknai arti dari kalimat.

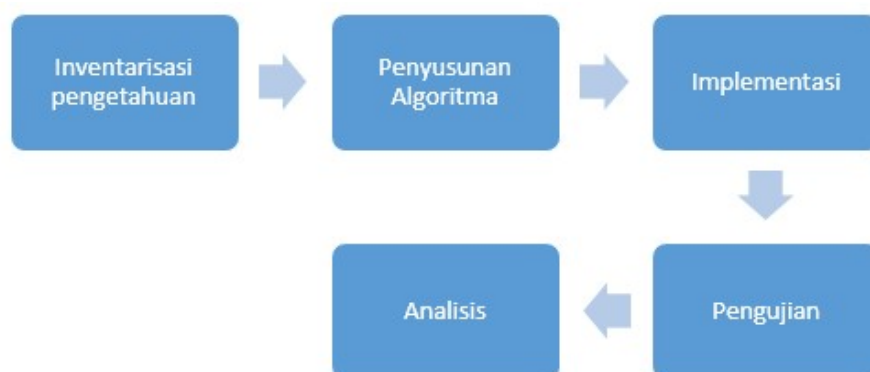
Dalam interaksi antara komputer dan manusia, terdapat kajian mengenai Pemrosesan bahasa alami atau *Natural Language Processing (NLP)*. Menurut Mooney (2008), *NLP* adalah cabang ilmu komputer dan linguistik yang mengkaji berbagai aspek pemrosesan bahasa alami sehingga dapat terjadi suatu interaksi antara manusia dan komputer dengan perantara bahasa alami. NLP memiliki peranan dalam mengurangi permasalahan ambiguitas yang ada pada teks berbahasa Indonesia. Tahapan dalam NLP yang berperan dalam menangani keambiguitasan ini adalah *Part-of-Speech (POS) Tagging*. *POS tagging* merupakan salah satu tahapan dari kajian NLP untuk menentukan kelas kata. Hasil *POS 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*.

Secara umum, terdapat dua aturan dalam *POS tagging*, yaitu Aturan leksikal yang digunakan untuk mencari tag yang paling sering digunakan dalam dokumen dan aturan kontekstual membantu menangani masalah ambiguitas. Penelitian mengenai *POS tagging* dalam konteks bahasa Indonesia sudah beberapa kali dilakukan, diantaranya adalah *Hidden Markov Models* dan *Rule based* untuk dokumen bahasa Indonesia (Widhiyanti Harjoko, 2012) dan *Brill Tagger* untuk dokumen bahasa Indonesia (Christanti, 2012). Dua penelitian tersebut belum memberi perhatian yang signifikan terhadap ambiguitas secara khusus dalam proses *POS tagging*.

Sementara itu, terdapat beberapa metode yang dapat digunakan dalam menyelesaikan permasalahan terkait mesin pembelajar, salah satunya adalah metode *rule based*. *Rule based* merupakan salah satu metode yang menggunakan aturan-aturan untuk menyajikan pengetahuannya. Metode ini menggunakan modus ponens (*jika-maka*) sebagai dasar untuk mengonfigurasi aturan (Lusiani, 2011). Metode ini cocok digunakan untuk menyelesaikan permasalahan yang memiliki lingkup terbatas dengan analisis yang mendalam.

## 2. METODE PENELITIAN

Penelitian ini tergolong ke dalam penelitian riset eksperimental. Menurut Jaedun (2011), metodologi riset berkaitan dengan pembahasan mengenai bagaimana cara memperoleh pengetahuan. Adapun desain implementasinya disusun sebagaimana dalam Gambar 1.



Gambar 1. Diagram Kerja Penelitian

## 2.1 Sumber Pengetahuan

Pengetahuan mengenai pola kalimat dalam bahasa Indonesia sebagai dasar penyusunan algoritma diperoleh dari buku-buku sintaksis. Beberapa buku yang menjadi rujukan adalah : *Pengantar Sintaksis Indonesia* (Wojowasito, 1976), *Dasar-Dasar Analisis Sintaksis* (Parera, 2009), *Sintaksis 2: Keselarasan Fungsi, Kategori, dan Peran Dalam Klausa* (Chaer, 2010), *Tatabentuk Bahasa Indonesia* (Muslich, 2009), *Jengjala Bahasa Indonesia* (Lubis, 1993), Psikolinguistik: Pengantar Pemahaman Bahasa Manusia (Dardjowidjojo, 2014), *Sintaksis Bahasa Indonesia: Pendekatan Proses* (Chaer, 2009).

## 2.2 Metode Analisis

Data awal yang digunakan pada penelitian ini adalah dokumen korpus Bahasa Indonesia. Korpus didapatkan dari pengembang aplikasi POS Tag Bahasa Indonesia dari Fakultas Ilmu Komputer Universitas Indonesia. Selanjutnya, data pengujian berupa kumpulan kata-kata ambigu diperoleh dari contoh penggunaan homonim dan polisemi. Metode yang digunakan untuk menganalisis data dalam penelitian ini adalah metode deskriptif. Metode deskriptif adalah teknik yang digunakan seorang peneliti untuk menggambarkan kenyataan situasi atau peristiwa dengan runtut dan sistematis (Panjaitan Iqbal, 2006). Metode ini digunakan untuk mengelola secara mendalam pengetahuan yang diperoleh dari literatur, penyusunan algoritma, dan data yang diujicobakan pada implementasi. Data-data tersebut diferivikasi dan disusun kembali secara sistematis sesuai dengan bingkai pemetaan masalah yang dikaji.

## 2.3 Batasan Masalah

Untuk menghindari dari pembahasan yang meluas, ada beberapa batasan pada penelitian ini, yaitu:

- 2.3.1 Penelitian ini berfokus pada algoritma penentuan kelas kata dari kata-kata bahasa Indonesia yang memiliki sifat ambigu, sehingga pemilihan metode dan *tool* POS tagging tidak menjadi pembahasan dalam penelitian ini. Artinya, metode dan alat apa pun yang akan digunakan dapat diasumsikan telah mumpuni dalam melakukan *POS tagging*.
- 2.3.2 Aspek ambiguitas yang menjadi cakupan dalam pembahasan penelitian ini adalah ambiguitas dalam tingkat leksikal, tidak mencakup ambiguitas dalam tingkat gramatikal dan fonetik, yakni meliputi homograf, homonim, dan polisemi.
- 2.3.3 Kata-kata ambigu yang menjadi objek penelitian ini berbentuk kata dasar, bukan berbentuk frasa atau kata berimbuhan.

## 3. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah sebuah algoritma, berupa alur pemikiran yang logis dan terbatas untuk menyelesaikan suatu pekerjaan atau suatu masalah dengan luaran yang benar. Algoritma inilah yang kemudian dijadikan landasan (pedoman) untuk membuat program komputer. Algoritma mempunyai aturan penulisan sendiri yang disebut dengan notasi algoritma. Notasi algoritma tidak tergantung dari spesifikasi bahasa pemrograman tertentu dan komputer yang mengeksekusinya. Notasi algoritma merupakan bahasa universal yang dapat diterima oleh semua bahasa pemrograman yang ada. Untuk membuat algoritma dari suatu permasalahan, biasanya digunakan salah satu dari tiga buah notasi algoritma yang dikenal, yaitu uraian kalimat deskriptif, *flowchart*, atau *pseudocode*.

### 3.1 Inventarisasi Pengetahuan

Rule based menjadikan aturan-aturan sebagai dasar pemilihan keputusan. Aturan-aturan yang digunakan dalam algoritma ini dihasilkan melalui pemahaman terhadap pola kalimat yang ada dalam bahasa Indonesia. Dari berbagai sumber referensi sebagaimana disebutkan sebelumnya, didapatkan beberapa pengetahuan, diantaranya adalah:

- a Bahasa Indonesia memiliki Pola Dasar Kalimat Inti. Pola dasar kalimat inti (PDKI) bahasa In-

donesia didasarkan pada struktur saling ketergantungan, distribusi, dan tata laku sintaksis antara kelas-kelas kata penyusunnya. Wujud paling kecil dari PDKI harus terdiri dari dua unsur kelas kata, saling bergantung, tetapi masing-masing memiliki kemungkinan untuk diperluas tanpa mengganggu antarmenghubung mereka dalam wujud yang paling kecil tadi. Semua kalimat bahasa Indonesia dalam wujudnya yang terkecil adalah PDKI. Kombinasi PDKI bahasa Indonesia adalah sebagai berikut:

- o NP + NP  
Bibi buruh
- o NP + AP  
Suara merdu
- o NP + VP  
Petani mengeluh
- o NP + VP + NP  
Kakak menggendong adik
- o NP + VP + NP + NP  
Ibu membelikan adik baju

Pola-pola dasar di atas, dapat diperluas selama tidak merubah pola PDKI, baik dalam frase maupun klausa. Beberapa contoh perluasan PDKI antara lain:

- o NP - AUX - VP  
Kamu masih menunggu
- o NP - M - VP - NP  
Saya bisa mendapatkan perhatian

- b Pola pengujian kelas kata dapat dilakukan dengan dua cara, yakni aspek morfologis dan sintaksis. Secara morfologis, sebuah kata akan ditinjau *adakah imbuhan dalam kata tersebut?*, jika ada, *bagaimana polanya?*. Sebagai contoh, Jika imbuhan berupa awalan ke- dan akhiran an, maka termasuk kata benda, seperti dalam kata kemampuan. Sedangkan aspek sintaksis menaruh perhatian terhadap kata yang mendahului dan mengikutinya. Misalnya Jika sebuah kata diikuti kata ini, maka kata tersebut termasuk kata benda, sebagaimana dalam kata kursi ini. Pola pengujian kedua inilah yang digunakan dalam menghasilkan algoritma penentuan kelas kata ini.
- c Terdapat tujuh tagset (label kelas kata) dalam Kamus Besar Bahasa Indonesia (KBBI) versi IV, yaitu: *n* untuk nomina, *v* untuk verba, *a* untuk ajektiva, *adv* untuk adverbial, *num* untuk numeralia, *p* untuk partikel, dan *pron* untuk pronomina.

### 3.2 Penyusunan Algoritma

Terdapat beberapa tahapan dalam penyusunan algoritma penentuan kelas kata ambigu bahasa Indonesia, sebagaimana dijelaskan berikut:

#### 3.2.1 Menentukan strategi penalaran

Seperti yang telah dijelaskan pada bab sebelumnya, terdapat dua pendekatan dalam penalaran *rule based*, yakni: *forward chaining* dan *backward chaining*. Pendekatan yang digunakan dalam penalaran penelitian ini adalah *backward chaining*. Alur kerja dari pendekatan ini adalah, diberikan sebuah calon target, kemudian aturan yang aksinya mengandung

target di-trigger. *Backward chaining* ini cocok untuk menelusuri fakta yang masih belum lengkap, disebut juga *goal driven*. Dalam kajian ilmu lain, penalaran ini memiliki esensi yang sama dengan pemahaman deduktif atau *bottom-up*.

### 3.2.2 Menghimpun kata-kata ambigu

Penelitian ini berfokus pada penentuan kelas kata pada kata-kata yang memiliki sifat ambigu pada tingkat leksikal. Oleh karenanya, penulis menghimpun beberapa kata yang termasuk homonim, homograf dan polisemi. Penulis mengumpulkan 70 kata ambigu yang berasal dari kamus polisemi dan KBBI. Selanjutnya, kata-kata tersebut dikembangkan menjadi kalimat sederhana yang akan digunakan dalam penyusunan algoritma dan pengujian. Beberapa contoh kata ambigu dan penggunaannya yang dihimpun adalah sebagai berikut:

- Tahu : Ayah makan *tahu* (n)  
           Ayah tidak *tahu* masalah itu (v)
- Bisa : Dia *bisa* melakukannya (v)  
           Aku terkena *bisa* ular (n)

### 3.2.3 Membuat pola keambiguitasan

Pada tahap ini, penulis memetakan ambiguitas dari setiap kata ambigu yang telah terhimpun. Sebagai contoh, dari dua contoh di atas dapat dikelompokkan ke dalam kata ambigu yang berjenis *v-n*. Untuk menelusuri label kelas kata yang lebih dari satu, penulis menggunakan acuan kelas kata yang terdapat pada KBBI. Pola ambiguitas hasil pemetaan disajikan dalam tabel 1.:

**Tabel 1.** Kombinasi Pola Keambiguitasan

No.	Kombinasi	Kata
1	n-v	<i>Apel, tahu, serang, serampang, gelar, karang, tabung, suling, tanggal, bisa, abuk, acah, adang, agan, ajak, ajun, alam, ambal, ambung, anggur, apel, arak, babat, badik, bangun, banting, barut, baur, belum</i>
2	a-n	<i>Genting, rapat, seminar, seksi, sedan, alang, alit, amfang, ampuh, arah, aring, asi, awang, badar, badung, bagal, banci, jamak, basi, baung, bayu</i>
3	a-v	Seret, ampai, mental
4	n-p	Dasar, bagi, guna
5	adv-n	Semi, barang, baku
6	a-n-v	Serak, seri, babak, alir, aras, kabur
7	adv-n-v	Selang, anggar
8	p-a-n	Asal, ala, bahana

### 3.2.4 Menyusun aturan

Dari pola ambiguitas yang didapat, selanjutnya penulis menyusun aturan penentuan kelas kata berdasarkan domain pengetahuan sintaksis. Sebagai gambaran, kata tahu memiliki dua kelas kata, yakni n dan v. Dalam aturan sintaksis, disebutkan bahwa ciri kalimat nominal adalah diikuti kata *ini/itu/tersebut*. Hal ini berarti jika kata tahu diikuti kata itu, maka tahu dalam konteks ini memiliki kelas kata nominal bukan verbal. Langkah ini menghasilkan 71 aturan yang dibagi dalam 4 kelompok.

- Aturan [1. ] digunakan untuk mendeteksi kata sebelum x

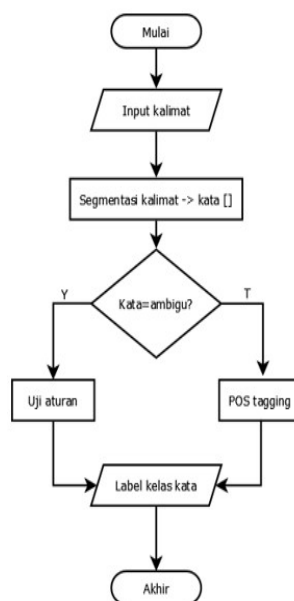
- Aturan [2. ] digunakan untuk mendeteksi label kata sebelum  $x$
- Aturan [3. ] digunakan untuk mendeteksi kata setelah  $x$
- Aturan [4. ] digunakan untuk mendeteksi label kata setelah  $x$

### 3.2.5 Notasi Algoritma

Hasil dari penelitian ini disajikan dalam bentuk kalimat deskripsi dan diagram alir (flowchart). Algoritma penentuan kelas kata pada POS tagging Bahasa Indonesia dideskripsikan sebagai berikut:

- Input kalimat yang akan dilakukan POS tagging.
- Segmentasi kalimat menjadi kata per kata
- Mengecek keambiguitasan kata dengan memeriksa apakah ada kata yang memiliki lebih dari satu kelas kata.
- Jika tidak, lakukan POS tagging. Jika ada, lakukan POS tag pada kata yang memiliki tepat satu kelas kata. Untuk kata yang memiliki lebih dari satu kelas kata, lakukan langkah selanjutnya.
- Uji dengan aturan-aturan yang sudah dibuat untuk menentukan kelas kata yang paling tepat.
- Tampilkan label kata hasil *POS tagging*.

Sedangkan penyajian algoritma penentuan kelas kata pada kata-kata yang memiliki sifat ambigu dalam POS tagging bahasa Indonesia dalam bentuk flowchart dapat dilihat pada gambar 2.



**Gambar 2.** Diagram Alir *flowchart* algoritma

### 3.3 Implementasi

Aturan-aturan keambiguitasan yang telah disusun kemudian dikombinasikan dengan algoritma POS tagging. Perlu dipahami kembali, bahwa algoritma ini bisa saja berbeda dengan algoritma lain bergantung dengan algoritma POS tagging yang digunakan. Seperti contoh, pada POS tagging dengan metode statistik ada tahapan analisis morfologi. Berbeda halnya dengan metode POS tagging yang digunakan dalam penelitian ini. POS tagging dalam penelitian ini mengacu pada korpus dari divisi bahasa fakultas ilmu komputer universitas Indonesia yang tidak membutuhkan adanya tahapan analisis morfologi. Oleh karena itu, pengembang juga perlu memahami karakteristik dari metode dan alat POS tagging yang digunakan.

Dalam pengimplementasian algoritma yang telah disusun, aplikasi POS tagging yang digunakan adalah aplikasi yang dikembangkan oleh kelompok belajar *NLP* UIN Sunan Kalijaga, dengan spesifikasi sebagai berikut:

DBMS : MySQL  
 Server localhost : XAMPP versi 3.2.2  
 Browser : UC Browser versi 6.1.2015.1007

Gambar 3. dan gambar 4. merupakan tampilan dari implementasi algoritma penentuan kelas kata ambigu.

Gambar 3. Tampilan aplikasi POS-tagging

Kata	:	dia	tahu	cara	membuat	tahu	yang	paling	gurih
Hasil Rule	:	pron	v	n	v	n	p	adv	a
Rule Ke	:	#	#2.7	#	#	#2.2	#	#	#
Database Korpus	:	pron	n	n	v	n	p	adv	a

Gambar 4. Tampilan Implementasi algoritma pada aplikasi *POS tagging*

Gambar 4. menunjukkan bahwa algoritma berhasil diterapkan untuk mendeteksi label kelas kata dari kata tahu yang memiliki dua kelas kata. Sedangkan kata yang hanya memiliki satu kelas kata, tidak menghasilkan nilai Rule ke- karena langsung dieksekusi sebelum ke pengujian aturan.

### 3.4 Pengujian

Pengujian terhadap algoritma ini dilakukan terhadap 70 kata ambigu yang telah dihimpun pada tahap penyusunan algoritma. Dari ke-70 kata itu, dikembangkan menjadi 100 kalimat sederhana untuk selanjutnya dilakukan proses POS tagging. Kalimat yang diujikan dihimpun dari Kamus Polisemi Bahasa Indonesia yang diterbitkan oleh Penerbit Andi pada 2004 dan contoh-contoh penggunaan homonim dan homograf dalam kalimat yang diakses di web pembelajaran sintaksis leksikal.

Hasil pengujian yang telah dilakukan terhadap aplikasi ditampilkan pada tabel 2:

**Tabel 2.** Hasil Pengujian

DATA	NILAI
Jumlah kata ambigu	70 kata
Jumlah kalimat yang diujikan	100 kalimat
Jumlah kata yang diproses	460 token
Jumlah pelabelan kata yang benar	360 kata
Jumlah pelabelan kata ambigu yang benar	92 kata
Jumlah pelabelan kata ambigu yang tidak sesuai	2 kata
Jumlah gagal aturan	6 kata

### 3.5 Analisis

Dari pengujian yang telah dilakukan, terdapat beberapa analisis yang diperoleh, diantaranya: Algoritma berhasil menangani 92 dari 100 kata ambigu (92%). Namun, algoritma penentuan kelas kata ini sangat dipengaruhi oleh ketersediaan aturan untuk menangani keambiguitasan kelas kata yang ada. Sebagai contoh, algoritma ini belum dapat menangani kata-kata yang tidak terdapat dalam aturan, seperti kata ambigu yang berada di awal kalimat (gambar 5) dan ketika terdapat dua kata ambigu yang berdampingan (Gambar 6). dua kata ambigu yang berdampingan akan bermasalah ketika kata pertama tidak memiliki label kelas kata yang benar. Hal ini mengakibatkan aturan yang diterapkan pada kata ambigu kedua menjadi tidak sesuai.



**Gambar 5.** Tampilan Algoritma bermasalah; kata ambigu berdampingan





Gambar 6. Tampilan Algoritma bermasalah; kata ambigu di awal kalimat

Pada gambar 5., kata ambigu seret yang memiliki kelas kata ajektiva dan verba terletak di awal kalimat. Penanganan kata ambigu yang berada di awal kalimat belum tertangani oleh algoritma. Pada Gambar 6., kata tahu (memiliki kelas kata noun dan verba) berdampingan dengan kata mental (memiliki kelas kata ajektiva dan verba), yang mana keduanya merupakan kata ambigu. Aturan yang terdapat dalam algoritma ini belum dapat menangani dua kata ambigu yang berdampingan. Selain itu, tagset KBBI yang digunakan dalam memberi label kelas kata kurang cocok untuk diterapkan dalam *POS tagging*. Hal ini dikarenakan KBBI memberi kelas kata yang masih global. Sebagai contoh, dalam KBBI semua kata kerja dilabeli dengan v, padahal pada pembagiannya terdapat kata kerja transitif dan intransitif. Masing-masing memiliki fungsi struktural yang berbeda. Boleh jadi, suatu kata yang berlabel v adalah transitif, sehingga berkaitan erat dengan kata yang mengikutinya sebagai objek. Objek sendiri memiliki bisa ditempati kata benda dan frasa kata benda. Oleh karenanya, dibutuhkan tagset label kelas kata yang dapat mencakup kebutuhan dalam proses *POS tagging*.

Selain itu, ketersediaan data pada korpus juga sangat mempengaruhi hasil dari implementasi algoritma ini. Jika suatu kata ambigu yang memiliki beberapa kelas kata tidak semua labelnya terdapat pada korpus, kata tersebut tidak akan dilabeli dengan benar. Terlebih lagi, jika kata tersebut sama sekali tidak ada pada korpus, maka hasil dari *POS tagging* tidak maksimal. Karena itu, dibutuhkan korpus data yang mencakup seluruh kata bahasa Indonesia beserta kelas katanya dengan lengkap.

#### 4. KESIMPULAN

Algoritma penentuan kelas kata ambigu pada *POS tagging* Bahasa Indonesia berhasil disusun dan diimplementasikan dengan menerapkan 71 aturan sintaksis yang digunakan dalam algoritma ini. Hasil dari pengujian menunjukkan bahwa algoritma ini berhasil memberikan label 92 kata dari 100 kata ambigu dengan benar, 2 kata tidak benar, dan 6 kata tidak tertangani oleh algoritma. Beberapa faktor yang mempengaruhi kinerja algoritma ini adalah kelengkapan aturan, label kelas kata dan korpus yang digunakan dalam proses *POS tagging*.

#### DAFTAR PUSTAKA

##### References

- Chaer, A., 2010. Sintaksis 2 Keselarasan Fungsi, Kategori Peran dalam Klausa. Universitas Muhammadiyah Surakarta, Surakarta.
- Christanti, V., 2012. Implementasi Brill tagger untuk Memberikan *POS tagging* pada Dokumen Bahasa Indonesia. Universitas Tarumanegara, Jakarta.
- Dardjowidjojo, S., 2014. Psikolinguistik. Keenam penyunt. Yayasan Obor Indonesia, Jakarta.
- Jaedun, A., 2011. Metodologi Penelitian Eksperimen. UNY, Yogyakarta.
- Lubis, H., 1993. Jengala Bahasa Indonesia. Angkasa, Bandung.

- Lusiani, 2011. Modul Sistem Pakar. STIKOM, Surabaya.
- Mooney, R., 2008. CS 388: Natural Language Processing Introduction. University of Texas, Austin.
- Muslich, M., 2009. Tata Bentuk Bahasa Indonesia. PT. Bumi Aksara, Jakarta.
- Na'im, A. Syaputra, H., 2011. Kewarganegaraan, Suku Bangsa, Agama dan Bahasa Sehari-hari Penduduk Indonesia: Hasil Sensus Penduduk 2010. Badan Pusat Statistik, Jakarta.
- Panjaitan Iqbal, D., 2006. Matinya Rating Televisi. Cetakan I penyunt. Yayasan Obor Indonesia, Jakarta.
- Parera, J., 2009. Dasar-dasar Analisis Sintaksis. Erlangga, Yogyakarta.
- Solehuddin, D., 2014. Bahasa Indonesia 10 Besar Dunia. [https://www.kompasiana.com/pengamatbijk/bahasa-indonesia-10-besar-dunia\\_52fef586ea834806d8b4598](https://www.kompasiana.com/pengamatbijk/bahasa-indonesia-10-besar-dunia_52fef586ea834806d8b4598), diakses 23 Desember 2017. Tim — Bahasa, 2008. *Kamus Besar Bahasa Indonesia*. Pusat Bahasa, Jakarta.
- Widhiyanti, K. Harjoko, A., 2012. POS tagging Bahasa Indonesia dengan HMM dan Rule based. UGM, Yogyakarta.
- Wojowasito, S., 1976. Pengantar Sintaksis Indonesia. Sintha Dharma, Bandung.
-