

Penerapan *Natural Language Processing* Untuk Mengidentifikasi Kalimat Ambigu pada Surat Kabar Menerapkan Metode *Shift Reduce Parsing*

Elya Wita

Prodi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: elyawitaaa@gmail.com

INFORMASI ARTIKEL

Article History

Received : Nov 15, 2021

Accepted : Nov 30, 2021

Published : Nov 30, 2021

KORESPONDENSI

Email: elyawitaaa@gmail.com

A B S T R A K

Natural Language Processing berfokus pada interaksi antara komputer dengan bahasa natural yang digunakan manusia dengan kata lain proses komputer melakukan ekstraksi informasi dari input berupa natural language dan/atau menghasilkan output yang juga berupa *natural language*. Surat kabar adalah lembaran tercetak yang memuat laporan yang terjadi di masyarakat. Namun Judul pada surat kabar seringkali mempunyai makna yang ambigu. Ketika kata, frasa, kalimat, atau bahkan paragraf yang ditulis atau diucapkan mengandung ambiguitas, secara otomatis tidak efektif dan komunikatif bagi pembaca atau pendengar. Untuk itu perlu dilakukan identifikasi kalimat agar memudahkan pembaca untuk memahami makna dari sebuah judul yang ambigu. Untuk memudahkan mengidentifikasi kalimat ambigu pada surat kabar membutuhkan algoritma *shift Reduce parsing*, dengan menggunakan algoritma *Shift Reduce parsing* dapat berguna untuk mengetahui maksud yang ingin disampaikan oleh penulis berita.

Kata Kunci: *Natural Language Processing*; Kalimat Ambigu; *Shift Reduce Parsing*

A B S T R A C T

Natural Language Processing focuses on the interaction between computers and the natural language used by humans. In other words, the computer process extracts information from input in the form of natural language and / or produces output which is also in the form of natural language. Newspapers are printed sheets containing reports that occur in the community. However, the titles of newspapers often have an ambiguous meaning. When a written or spoken word, phrase, sentence, or even paragraph contains ambiguity, it is automatically ineffective and communicative for the reader or listener. For this reason, it is necessary to identify sentences to make it easier for readers to understand the meaning of an ambiguous title. To make it easier to identify ambiguous sentences in newspapers requires a *shift Reduce parsing* algorithm, using the *Shift Reduce parsing* algorithm can be useful to find out what the news writer wants to convey.

Keywords: *Natural Language Processing*; Ambiguous Sentences; *Shift Reduce Parsing*

1. PENDAHULUAN

Surat kabar adalah media massa yang populer kalangan masyarakat. Baik itu tingkat atas, maupun tingkat bawah. Hal ini disebabkan karena surat kabar mempunyai kelebihan yaitu dapat dibaca kapan saja dan informasi yang diberikan lebih rinci dan detail, juga harganya yang relatif terjangkau jika dibandingkan dengan media massa lainnya.

Informasi dari seluruh wilayah maupun penjuru dunia dapat diketahui melalui surat kabar. Namun, tanpa disadari tulisan atau judul yang terdapat pada surat kabar terkadang mengandung makna ambigu. Ini dapat menjadi masalah dikarenakan penerima informasi sulit untuk memahami maksud dari informasi tersebut karena terdapat kata-kata atau kalimat yang mempunyai makna ganda. Sebagian orang salah mengartikan apa yang dibaca dan yang didengarnya. Hal ini terjadi baik dalam secara lisan maupun tulisan. Pembaca harus jeli dalam menangkap makna yang dimaksudkan oleh penulis tersebut karena mempunyai makna yang ganda hingga dapat menimbulkan keraguan dan bingung dalam menafsirkan makna yang dimaksudkan oleh pemberi informasi.

Ambiguitas dalam berbahasa dapat terjadi pada pihak pembaca atau pendengar. Tidak sedikit bahasa di media massa menggunakan bahasa yang ambigu. Ketika kata, frasa, kalimat, atau bahkan paragraf yang ditulis atau

diucapkan mengandung ambiguitas, secara otomatis tidak efektif dan komunikatif bagi pembaca atau pendengar. Penulisan judul yang ambigu terkadang disengaja agar menarik minat pembaca. Namun hal tersebut sangat berdampak negatif karena Keambiguan dapat mengakibatkan terjadinya lebih dari satu makna pada saat pembicaraan lisan ataupun dalam keadaan tertulis. Saat pembicaraan lisan mungkin dapat diantisipasi dengan dilafalkan perlahan, sedangkan untuk yang tertulis apabila terdapat kekurangan tanda baca maka makna suatu kata atau kalimat menjadi berbeda dari makna yang diinginkan oleh penulis..

Natural Language Processing (Pemrograman Bahasa Alami) merupakan pembuatan program yang memiliki kemampuan untuk memahami bahasamanusia. Dalam kaitan interaksi antara komputer dan bahasa manusia, *Natural Language Processing* (NLP) berperan untuk mengurangi permasalahan ambiguitas yang ada pada teks berbahasa Indonesia seperti yang ada pada judul surat kabar. Terdapat beberapa metode pada penanganan ambiguitas diantaranya *Shift-Reduce Parsing*.

Penelitian terdahulu mengenai kalimat ambigu pernah dilakukan oleh Fachrurrozi, Yusliani dan Agustin menggunakan *Shift Reduce Parsing* untuk mengetahui ambiguitas kalimat bahasa indonesia dengan tingkat akurasi diperoleh sebesar 83% [1]. Penelitian menggunakan metode *Shift Reduce Parsing* juga pernah dilakukan oleh Yusiani, Yunita dan Octaviani penelitian ini bertujuan untuk mengenai frasa benda pada kalimat, hasil dari penelitian ini membuktikan bahwa metode *Shift reduce parsing* dapat digunakan untuk mengidentifikasi pola frasa benda pada kalimat.

2. METODOLOGI PENELITIAN

2.1 Natural Language Processing

Natural Language Processing (NLP) merupakan salah satu cabang ilmu AI yang fokus pada pengolahan bahasa natural. Bahasa natural merupakan bahasa yang umum digunakan manusia dalam berkomunikasi. Fokus NLP adalah interaksi antara komputer dengan bahasa natural yang digunakan manusia dengan kata lain proses komputer melakukan ekstraksi informasi dari input berupa *naturallanguage* dan/atau menghasilkan output yang juga berupa *natural language* [2-3].

2.2 Part of Speech Tagging

Part of speech tagging (*pos tagging*) adalah pemberian label pada kata. Fungsi label tersebut adalah untuk menunjukkan fungsi kata di dalam kalimat. Macam – macam pelabelan tersebut antara lain kata sifat, kata benda, kata kerja dan lain-lain. Kumpulan label atau *tag part of speech* atau kelas kata disebut sebagai tagset. Dalam membangun *POS-Tagger* harus memperhatikan struktur kalimat dari korpus anotasi sebagai acuan dalam memberikan tag pada data pengujian. pemberian tag menggunakan *Hidden Markov Model-Viterbi* (HMM- *Viterbi*). Korpus digunakan dalam *POS-Tagger HMM-Viterbi* dibuat menggunakan tagset yang dibangun oleh Fam Rashel,dkk [4-5].

2.3 Kalimat Ambigu

Kalimat ambigu merupakan kalimat yang mempunyai dua arti atau lebih.kalimat ambigu dapat menghambat kelancaran berkomunikasi karena pihak penerima informasi dapat menafsirkan arti yang berbeda dengan arti yang dimaksud oleh pembicara atau penulis [6-7]. Suatu kalimat dikatakan ambigu apabila kalimat tersebut memiliki lebih dari satu tafsir. Makna suatu kalimat ditentukan oleh fungsi sintaksis deretan kata pembentuknya. Artinya,dalam kalimat yang ambigu, suatu deretan kata dapatmemiliki fungsi sintaksis lebih dari satu. Akibatnya, pola kalimat yang terbentuk pun lebih dari satu pula.

3. HASIL DAN PEMBAHASAN

Natural Language Processing (NLP) telah dikenal beberapa metode yang sering digunakan untuk menangani masalah ambiguitas. Salah satu metode yang dapat digunakan dalam mengenali struktur tata bahasa suatu kalimat adalah *Shift Reduce Parsing*.

Pada penelitian ini, untuk mengidentifikasi kalimat ambigu pada surat kabar dibagi menjadi beberapa bagian yaitu preprocessing, identifikasi, Pos-tagging dan implementasi algoritma. Pada tahap *preprocessing*, dilakukan proses *case folding* untuk mengubah kalimat menjadi huruf kecil kemudian dilakukan proses *tokenizing* untuk memecah kalimat menjadi kata per kata. Kemudian identifikasi apakah struktur kalimat pada kalimat masukan ambigu atau tidak. Proses selanjutnya yaitu *Part-Of-Speech Tagging* sistem akan melakukan proses pelabelan kata pada kata per kata.

3.1 Penerapan Part-Of-Speech Tagging

Untuk mengidentifikasi judul berita ambigu pada surat kabar, penulis melakukan identifikasi pada judul surat kabar yang mempunyai makna ambigu. Untuk memudahkan analisa maka penulis menentukan judul berita “Kematian ibu Alarm turunnya layanan kesehatan” adapun judul tersebut diambil dari surat kabarkompas. Untuk mengetahui makna sebenarnya dari judul berita tersebut penulis menggunakan metode *Shift Reduce parsing* untuk mengurai kalimat tersebut.

Pada tahap ini proses yang akan dilakukan adalah mengolah data masukan yang masih berupa data belum terstruktur menjadi data yang terstruktur, proses initerdiri dari tahapan yaitu *Case Folding* dan *tokenizing*.

1. Case folding

Proses *Case Folding* merupakan tahapan yang pertama yang dilakukan padaproses *Preprocessing*, pada tahap ini seluruh huruf dalam mulai dari „a“ sampai „z“ pada sumber teks diubah menjadi huruf kecil, kemudian karakter selain huruf dihilangkan dan dianggap sebagai pembatas. Namun tidak semua dokumen konsisten dengan penggunaan huruf capital, maka dari itu *Case Folding* mengkonversi keseluruhan teks dalam menjadi huruf kecil. Hasil *case folding* berdasarkan data masukan dapat dilihat dari tabel 1 dibawah ini:

Tabel 1. Hasil *case folding*

Kalimat	Hasil <i>case folding</i>
Kematian Ibu Alarm Turunnya Layanan Kesehatan.	kematian ibu alarm turunnya layanan kesehatan

2. Tokenizing

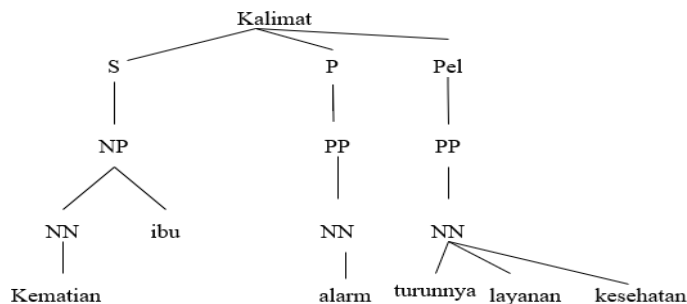
Tokenizing adalah proses pemotongan *string input* berdasarkan tiap katayang menyusunnya. Proses ini dilakukan dengan menghilangkan tanda baca dan memisahkannya per spasi. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua *token* ke bentuk huruf kecil Hasil *tokenizing* dapat dilihat pada tabel 2.

Tabel 2. Hasil Tokenizing6

Sebelum	Sesudah <i>tokenizing</i>
kematian ibu alarm turunnya kesehatan	Kematian ibu alarm turunnya layanan kesehatan

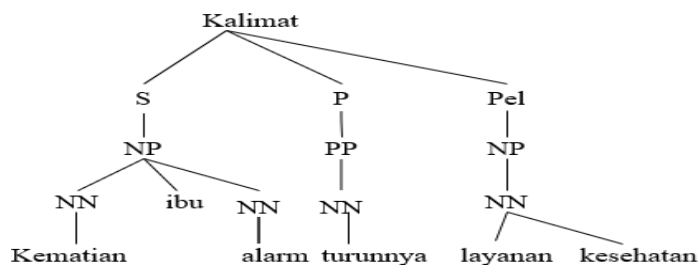
3. Part of Speech Tagging

Part of speech tagging (pos tagging) adalah pemberian label pada kata.Fungsi label tersebut adalah untuk menunjukkan fungsi kata di dalam kalimat. Label atau tag yang diberikan ke suatu kata dalam suatu kalimat menunjukkankelas kata (*word class*) dari kata tersebut. Hasil POS-Tagging dapat dilihat pada contoh dibawah ini. Kematian/ NN ibu/NP alarm/NNP turunnya/ NN layanan/NN kesehatan/NNP. Proses ini untuk mengenali apakah kalimat masukan tersebut merupakan kalimat ambigu atau tidak. Hasil dari identifikasi kalimat dapat dilihat pada gambar 1 berikut.



Gambar 1. Type-1 parsing sentence

Dari gambar 1 dapat diketahui makna dari kalimat “kematian ibu alarm turunnyaangka kesehatan” adalah kematian pada ibu, menjadi peringatan bahwa pelayanan kesehatan menurun. Sementara hasil dari identifikasi kalimat ambigu pada gambar 3.5 diketahui makna dari kalimat masukan tersebut adalah : Kematian seorang ibu bernama Alarm dikaitkan dengan menurunnya layanan kesehatan.



Gambar 2. Type-2 Parsing sentence

4. Algoritma Shift Reduce Parsing

Algoritma *Shift Reduce Parsing* merupakan tahap penyelesaian untukmelakukan identifikasi kalimat ambigu pada surat kabar. Dari pemrosesan sebelumnya didapat hasil sebagai berikut :kematian/NN = n ibu/NP = n alarm/NNP = n turunnya/NN = v layanan/NN = n kesehatan/NNP = n

Identifikasi kalimat ambigu pada surat kabar menggunakan algoritma shiftreduce parsing adalah sebagai berikut:

Tabel 3. Hasil Shift Reduce Parsing

LHS	RHS (input)	Aksi
[]	Kematian ibu alarm turunny layan kesehatan	
[kematian]	ibu alarm turunny layan	Shift kata ke LHS
[n ibu]	kesehatan turunny layan	shift
[n]	kesehatan turunny layan	reduce
[n alarm]	kesehatan turunny layan	shift
[n]	turunny layan kesehat	reduce
[n turunny]	layan kesehat	Shift
[n v]	layan kesehat	reduce
[s v]	Layan kesehat	reduce
[s v layan]	kesehat	Shift
[s v layan kesehat]	[]	shift
[s v n kesehat]	[]	reduce
[s v n kesehat]	[]	reduce
[s v n n]	[]	reduce
[s p n n]	[]	reduce
[s p n n]	[]	reduce
[s p n]	[]	reduce
[s p pel]	[]	reduce
[sentence]	[]	accept

Meskipun operasi utama *Parser* adalah *Shift* dan *Reduce*, sebenarnya ada empat kemungkinan tindakan yang dapat dilakukan oleh *Shift Reduce parsing*:

1. Dalam aksi *Shift*, simbol *input* berikutnya digeser ke atas tumpukan.
2. Dalam aksi *Reduce*, *Parser* mengetahui bahwa ujung kanan *handle* berada di sekitar bagian atas *Stack*. *Stack* kemudian harus menemukan ujung kiri *handle* di dalam *Stack* dan memutuskan dengan non-terminal apa yang akan menggantikan *handle*.
3. *accept*, *Parser* menyelesaikan penguraian dengan sukses.
4. *error*, pengurai menemukan bahwa kesalahan sintaks telah terjadi dan memanggil pemulihan kesalahan.

4. KESIMPULAN

Berdasarkan hasil dilakukan oleh penulis dari bab-bab sebelumnya diambil suatu kesimpulan yang merupakan hasil akhir dari penulisan yaitu sebagaiberikut: Kalimat ambigu dapat diidentifikasi dengan proses parsing (*Parsing sentence*). *Natural language processing* dapat diterapkan untuk mengidentifikasi kalimat ambigu dengan menggunakan *metode shift reduce parsing*. Ketiadaan tanda baca dalam kalimat yang digunakan sebagai masukan memperbesar peluang terjadinya kalimat yang ambigu.

REFERENCES

- [1] O. Anton, W. Prihartono, and S. Sos, "Surat Kabar & Konvergensi Media (Studi Deskriptif Kualitatif Model Konvergensi Media Pada Solopos)," vol.4, no. 1, pp. 105–116, 2016, doi: 10.12928/channel.v4i1.4210.
- [2] T. D. Rizki and N. Yusliani, "Rancang Bangun Sistem Pengecekan Ambiguitas Kalimat Berbahasa Indonesia Menggunakan Harmony Search Algorithm," *Annu. Res. Semin.*, vol. vol.2, NO., no. 1, pp. 173–176, 2016.
- [3] N. K. Wangsanegara and B. Subaeki, "Implementasi Natural Language Processing Dalam Pengukuran Ketepatan Ejaan Yang Disempurnakan (Eyd) Pada Abstrak Skripsi Menggunakan Algoritma Fuzzy Logic," *J. Tek. Inform.*, vol. 8, no. 2, 2015, doi: 10.15408/jti.v8i2.3185.
- [4] M. Fachrurrozi, N. Yusliani, and M. M. Agustin, "Identification of Ambiguous Sentence Pattern in Indonesian Using Shift-Reduce Parsing," *Proceeding 1st Int. Con ference Comput. Sci. Eng.*, pp. 61–63, 2014.
- [5] B. Widodo and S. Derwin, "Artificial Intelligence konsep dan penerapannya," in *Artificial Intelligence konsep dan penerapannya*, Seno, Ed. Yogyakarta: Andi, 2014, pp. 2–3.
- [6] H. Li, L. Tiung, S. Tetsuya, and W. Kam-fai, "Information Retrieval Technology," in *Information Retrieval Technology*, china: Springer berlin Heidelberg, 2008, pp. 71–82.
- [7] T. A. Kurniawan, "Pemodelan Use Case (UML): Evaluasi Terhadap beberapa Kesalahan dalam Praktik," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 1, p. 77, 2018, doi: 10.25126/jtiik.201851610.