

IMPLEMENTASI POS TAGGING DAN ALGORITMA ANTLR PARSER DALAM MEMERIKSA STRUKTUR KALIMAT BAHASA INDONESIA

Marandina Putri Prapasha, Husni Thamrin
Program Studi Teknik Informatika, Fakultas Komunikasi Dan
Informatika, Universitas Muhammadiyah Surakarta

Abstrak

Komunikasi merupakan langkah dimana pesan atau informasi disampaikan dari satu pihak ke pihak lainnya agar pesan tersebut dapat dipahami dengan baik. Komunikasi merupakan suatu aspek yang sangat penting bagi manusia, tidak hanya dalam ranah kehidupan berorganisasi, melainkan juga dalam kehidupan sehari-hari. Komunikasi akan berjalan dengan lancar jika pembicara menggunakan bahasa yang efektif, terstruktur, dan familiar bagi lawan bicara. Bahasa yang efektif merujuk pada penggunaan kata-kata yang mengikuti tata bahasa atau kaidah bahasa dengan baik dan benar, seperti struktur atau pola kalimat yang sesuai. Namun, pada umumnya banyak orang kurang memperhatikan struktur kalimat yang baik dan benar dalam menyampaikan pesannya kepada orang lain. Pada akhirnya, hal tersebut memiliki potensi untuk menimbulkan kesalahan dalam penggunaan bahasa. Tujuan dari penelitian ini adalah melakukan pemeriksaan terhadap struktur kalimat dengan cara mengidentifikasi subjek, predikat, objek, dan keterangan (SPOK) secara otomatis. Penelitian bermanfaat untuk mempermudah pemeriksaan struktur kalimat. Metode yang digunakan adalah penerapan algoritma Flair untuk *POS (Part Of Speech) Tagging* serta algoritma *ANTLR (ANother Tool Language Recognition)* untuk melakukan parsing terhadap struktur kalimat. *POS (Part Of Speech) Tagging* digunakan untuk mengelompokkan kata-kata ke dalam kategori tertentu seperti kata benda (noun), kata kerja (verb), dan kata sifat (adjective). Algoritma *ANTLR (ANother Tool Language Recognition)* digunakan untuk menentukan posisi kata atau frase sebagai SPOK dengan menggunakan grammar dan lexer untuk mengenali token yang dihasilkan dari proses *POS (Part Of Speech) Tagging*. Hasil penelitian ini adalah terdapat 424 kalimat yang benar dari 500 kalimat dan menghasilkan tingkat keberhasilan (%) sebesar 84,8%. Saran untuk penelitian mendatang adalah mengambil pertimbangan terhadap penggunaan perpustakaan NLP yang lebih lengkap dan lebih akurat, serta korpus yang lebih luas daripada perpustakaan Flair NLP yang digunakan dalam penelitian ini, untuk menetapkan label jenis kata. Selain itu, disarankan untuk mempertimbangkan perluasan struktur kalimat yang digunakan, tidak hanya terbatas pada kalimat tunggal, tetapi juga meliputi jenis kalimat lain seperti kalimat majemuk dan jenis kalimat lainnya, sehingga struktur kalimat dapat diperiksa dan diidentifikasi dengan lebih baik.

Kata Kunci: Struktur Kalimat, Parsing, *POS Tagging*, *ANTLR Parser*

Abstract

Communication is a step in which a message or information is conveyed from one party to another so that the message can be understood properly.

Communication is a very important aspect for humans, not only in the realm of organizational life, but also in everyday life. Communication will run smoothly if the speaker uses language that is effective, structured and familiar to the person speaking. Effective language refers to the use of words that follow grammar or language rules properly and correctly, such as appropriate sentence structures or patterns. However, in general, many people do not pay attention to good and correct sentence structure in conveying their messages to other people. In the end, this has the potential to cause errors in language use. The aim of this research is to examine sentence structure by automatically identifying subjects, predicates, objects and adverbs (SPOK). Research is useful for making it easier to examine sentence structure. The method used is the application of the Flair algorithm for POS (Part Of Speech) Tagging and the ANTLR (ANother Tool Language Recognition) algorithm for parsing sentence structures. POS (Part Of Speech) Tagging is used to group words into certain categories such as nouns, verbs and adjectives. The ANTLR (ANother Tool Language Recognition) algorithm is used to determine the position of words or phrases as SPOK by using grammar and lexer to recognize tokens produced from the POS (Part Of Speech) Tagging process. The results of this research were that there were 424 correct sentences out of 500 sentences and produced a success rate (%) of 84.8%. A suggestion for future research is to take into consideration the use of a more complete and more accurate NLP library, as well as a wider corpus than the Flair NLP library used in this study, to determine word type labels. In addition, it is recommended to consider expanding the sentence structures used, not only limited to single sentences, but also include other types of sentences such as compound sentences and other types of sentences, so that sentence structures can be better examined and identified.

Keywords: Sentence Structure, Parse, POS Tagging, ANTLR Parser

1. PENDAHULUAN

Bahasa merupakan sarana untuk mengkomunikasikan pikiran seseorang agar diketahui oleh orang lain. Kegiatan berkomunikasi memiliki peranan yang sangat signifikan bagi manusia, tidak hanya dalam konteks kehidupan berorganisasi, namun juga dalam kehidupan sehari-hari (Pohan and Fitria 2021). Komunikasi tidak hanya terbatas pada bentuk lisan, melainkan juga dapat dilakukan dalam bentuk tertulis. Komunikasi akan berjalan dengan lancar jika pembicara menggunakan bahasa yang efektif, terstruktur, dan familiar bagi lawan bicara (Fauzia 2018) (Ludiani, Dr. Laili Rahmawati, S.Pd., and Dr. Atiqa Sabardila 2022). Bahasa yang efektif merujuk pada penggunaan kata-kata yang mengikuti tata bahasa atau kaidah bahasa dengan baik dan benar, seperti struktur atau pola kalimat yang sesuai (Mizkat 2019). Ahli bahasa David A. Wilkins (Pratama, Kusumadewi, and Hidayat 2017) menyatakan, "tanpa tata bahasa anda hanya dapat menyampaikan sedikit hal; tanpa kosakata anda tidak dapat menyampaikan apa pun."

Pandangan ini menunjukkan bahwa ketika menyampaikan ide, pendapat, dan perasaan melalui tulisan kepada orang lain, penting untuk memperhatikan keakuratan dalam menggunakan bahasa, kosa kata, tata bahasa, dan ejaan (Martha and Situmorang 2018). Pada umumnya banyak orang kurang memperhatikan tata bahasa dalam menyampaikan bahasa yang efektif (Mizkat 2019). Pada akhirnya, hal tersebut memiliki potensi untuk menimbulkan kesalahan dalam penggunaan bahasa (Izzati 2016).

Penentuan pola kalimat biasa dilakukan secara analitis oleh orang yang terlatih. Namun untuk kebutuhan tertentu diperlukan penentuan secara otomatis menggunakan mesin. Contoh penerapannya adalah proses penerjemahan dan proses peringkasan teks secara otomatis (Abdiansah and Muhammad Qurhanul Rizqie 2023) (Sa, Kumar, and Kasmin 2023). Pada proses penerjemahan misalnya, perlu diperhatikan mana kata yang menjadi subjek atau predikat agar hasil terjemahan menjadi sesuai (Haryanti et al. 2020). Maka, dalam penelitian ini, peneliti akan melakukan pengecekan secara otomatis terhadap struktur kalimat, termasuk subjek, predikat, objek, dan keterangan (SPOK). Penentuan unsur pola kalimat dari sebuah teks disebut dengan *parsing*.

Beberapa penelitian telah dilakukan terkait pengenalan pola kalimat. Penelitian “Analisis Morfologi untuk Menangani *Out-of-Vocabulary Words* pada *Part-of-Speech Tagger* Bahasa Indonesia Menggunakan Hidden Markov Model” yang dilakukan oleh Febyana Ramadhanti, Yudi Wibisono, dan Rosa Ariani Sukamto pada tahun 2019 (Ramadhanti, Wibisono, and Sukamto 2019) menyatakan bahwa penggunaan metode AM pada sistem *POS (Part Of Speech) Tagger* dengan HMM memiliki tingkat ketepatan yang lebih tinggi daripada Pos Tagger tanpa AM dalam mengatasi kata-kata yang tidak dikenali (OOV).

Penelitian berjudul “Penentuan Kelas Kata Pada *Part of Speech Tagging* Kata Ambigu Bahasa Indonesia” oleh Ahmad Subhan Yazid dan Agung Fatwanto pada tahun 2018 (Yazid and Fatwanto 2018) menyimpulkan bahwa algoritma yang digunakan berhasil memberikan label yang tepat untuk 92 dari 100 kata ambigu, namun terdapat 2 kata yang diberi label secara tidak akurat, dan 6 kata tidak diidentifikasi oleh algoritma. Beberapa faktor yang mempengaruhi kinerja algoritma ini mencakup keakuratan aturan, label kelas kata, dan korpus yang digunakan dalam proses *POS (Part Of Speech) Tagging*.

Penelitian berjudul “An empirical evaluation of Lex/Yacc and *ANTLR (Another Tool Language Recognition) Parser generation tools*” yang disusun oleh Fransisco Ortin, Jose Quirog, Oscar Rodriguez-Prieto, dan Miguel Garcia pada tahun 2022 (Ortin et al. 2022) menunjukkan bahwa, untuk implementasi bahasa pemrograman dengan kompleksitas sedang oleh mahasiswa tahun ke- 3 jurusan Rekayasa Perangkat Lunak, alat *ANTLR (Another Tool Language Recognition)* menunjukkan manfaat yang signifikan dibandingkan dengan Lex/Yacc. Penggunaan *ANTLR (Another Tool Language Recognition)* meningkatkan tingkat kelulusan, persentase kehadiran ujian, nilai lab, dan nilai akhir.

Penelitian berjudul “Detection and Classification of Cross-language Code Clone Types by Filtering the Nodes of *ANTLR (Another Tool Language Recognition)*-generated Parse Tree” yang disusun oleh Sanjay B. Ankali dan Latha Parthiban pada tahun 2021 (Ankali and Parthiban 2021) menunjukkan bahwa deteksi klon lintas bahasa yang diusulkan melakukan teknik yang ada dengan menunjukkan presisi tertinggi 95,37% dalam menemukan semua jenis klon (1, 2, 3 dan 4) untuk 16.032 pasangan klon kode C dan CPP yang serupa secara semantik dari sanfoundry.com. Karena tata bahasa *ANTLR (Another Tool Language Recognition)* tersedia secara gratis, metodologi yang diusulkan dapat dikembangkan untuk menyertakan bahasa pemrograman lain guna membangun sistem rekomendasi metode yang kuat dan lebih andal untuk membantu proses forking perangkat lunak.

Penelitian berjudul “Aspek Pedagogik Implementasi Translator Notasi Algoritmik Berbasis Parsing LL(*) dan String Template” yang disusun oleh Wijanarto dan Ajib Susanto pada tahun 2014 (Wijanarto and Susanto 2014) menyatakan bahwa penggunaan model notasi algoritmik dengan LL(*) parsing dan template string menggunakan grammar *ANTLR (Another Tool Language Recognition)* telah berhasil dikembangkan menjadi sebuah alat pembelajaran dasar pemrograman yang dinamakan ETNA.

Penelitian berjudul “Penerapan Abstract Syntax Tree dan Algoritme Damerau-Levenshtein Distance untuk Mendeteksi Plagiarisme pada Berkas Source Code” yang ditulis oleh Stephanie Rusdianto dan Ria Chaniago pada tahun 2019 (Rusdianto and Chaniago 2018) menunjukkan bahwa tingkat akurasi metode pohon sintaks abstrak dan algoritma jarak Damerau-Levenshtein dalam mendeteksi plagiarisme, dengan

menggunakan aturan lexer dan parser grammar, mencapai 97,3076923%. Aplikasi memerlukan waktu komputasi selama 118,115 detik. Akurasi metode pohon sintaks abstrak dan Algoritma Jarak Damerau-Levenshtein dalam mendeteksi plagiarisme, hanya dengan menggunakan aturan lexer grammar, adalah 96,923076%. Waktu komputasi aplikasi untuk ini adalah 6,057 detik.

Penelitian berjudul “Model Translator Notasi Algoritmik ke Bahasa C” yang disusun oleh Wijanarto dan Achmad Wahid Kurniawan pada tahun 2012 (Dan and Intelijen 2012) menunjukkan bahwa hasil eksperimen dan pengujian terhadap teknik penulisan notasi dengan menggunakan grammar dari *ANTLR (Another Tool Language Recognition) Parser* untuk menyelesaikan masalah mencapai hasil yang diharapkan, yakni lebih cepat, lebih mudah dibaca (oleh orang Indonesia yang memiliki pemahaman terbatas tentang bahasa Inggris), dibandingkan dengan mengekspresikan langsung dalam bentuk bahasa (biasanya bahasa Inggris murni).

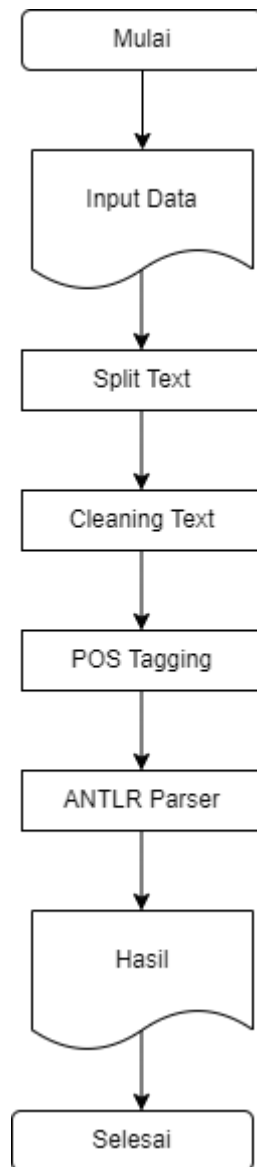
Sedangkan penelitian berjudul “Analisis dan Perancangan Domain Specific Language untuk Data Generator pada Relational Database” yang ditulis oleh Vania Natali dan Pascal Alfadian pada tahun 2019 (Natali and Alfadian 2019) menyebutkan bahwa tata bahasa yang dihasilkan *ANTLR (Another Tool Language Recognition)* bisa digunakan untuk melakukan analisis sintaksis pada DDL yang digunakan sebagai masukan untuk alat pembuat data yang telah dibangun.

Penelitian yang diusulkan ini berupaya menentukan pola kalimat SPOK (Subjek, Predikat, Objek, dan Keterangan) secara otomatis. Batasan dari penelitian ini yaitu hanya menggunakan kalimat tunggal. Perbedaan utama antara penelitian yang diusulkan dengan penelitian sebelumnya adalah penggabungan antara *POS (Part Of Speech) Tagging* dengan *ANTLR (Another Tool Language Recognition) Parser*. *POS (Part Of Speech) Tagging* berfungsi untuk memilah kata-kata dalam sebuah kalimat dan mengkategorikan setiap kata ke dalam jenis kata yang sesuai. Algoritma *ANTLR (Another Tool Language Recognition) Parser* digunakan untuk menentukan kata tersebut termasuk ke dalam Subjek, Predikat, Objek, atau Keterangan.

2. METODE

Pemeriksaan struktur atau pola kalimat ini menggunakan metode algoritma *ANTLR (Another Tool Language Recognition) Parser* dengan *POS (Part Of Speech) Tagging*.

Algoritma *ANTLR (Another Tool Language Recognition) Parser* merupakan suatu perangkat yang berguna untuk membaca, memproses, serta mengeksekusi atau menerjemahkan teks yang terstruktur (Natali and Alfadian 2019). *ANTLR (Another Tool Language Recognition) Parser* memanfaatkan modul *parser-jvm* dan *parser-js* supaya benar-benar berfungsi dalam produksi. Algoritma ini berfokus pada penentuan SPOK untuk setiap kata yang membentuk sebuah kalimat. *POS (Part-of-Speech) Tagging* merupakan proses mengklasifikasikan teks ke dalam kategori kata tertentu, seperti kata benda (noun), kata kerja (verb), atau kata sifat (adjective), yang bergantung pada makna dan konteksnya. *POS (Part Of Speech) Tagging* yang dilakukan dengan menggunakan modul NLTK. Modul tersebut merupakan kolaborasi komunitas programmer Python. Penggunaan algoritma *ANTLR (Another Tool Language Recognition) Parser* dengan *POS (Part Of Speech) Tagging* dapat mempermudah dalam menentukan subjek, predikat, objek, dan keterangan berdasarkan dari kelas kata. Sumber data yang digunakan berasal dari internet yaitu dibantu dengan Google dan ChatGPT. Selain itu, data teks 500 kalimat juga dikembangkan secara manual. Data yang digunakan untuk proses pemeriksaan ini yaitu sebuah file berupa .txt yang berisikan 500 kalimat tunggal. Untuk melakukan pemeriksaan struktur kalimat otomatis ini menggunakan bahasa pemrograman Python pada Google Colaboratory. Peneliti membuat loop parsing dan menghitung keberhasilan parsing dengan cara menghitung jumlah kalimat yang berstruktur dibagi dengan jumlah kalimat keseluruhan. Untuk mendapatkan hasil tingkat keberhasilan (%), peneliti juga melakukan parsing manual sebagai kontrol dengan tujuan untuk membandingkan hasilnya dengan hasil parsing yang dilakukan oleh komputer. Gambar 1. Menunjukkan proses pemeriksaan ini.



Gambar 1. Diagram Pemeriksaan Struktur Kalimat

Tahapan-tahapan yang dilakukan dengan metode *ANTLR (ANother Tool Language Recognition) Parser* antara lain :

2.1 Input Data

Menginputkan data merupakan tahap pertama dari proses pemeriksaan struktur kalimat secara otomatis. Pada tahap ini yang perlu dipersiapkan ialah data yang berisi 500 kalimat tunggal. Data yang disiapkan tersedia dalam format dokumen teks (.txt). Pada tahap ini, yang diperlukan adalah memasukkan atau menambahkan data yang akan diperiksa. Data didapatkan dari internet yaitu dibantu dengan Google dan

ChatGPT.

2.2 Split Text

Split data merupakan tahap kedua dari proses pemeriksaan struktur kalimat secara otomatis. Split text merupakan sebuah istilah yang digunakan untuk dilakukannya proses pemisahan kalimat utuh menjadi beberapa kata yang terpisah. Tujuan dari split text ini ialah untuk memisahkan antar kata dalam sebuah kalimat agar lebih mudah untuk dipahami. Pada tahap ini yang dilakukan ialah memisahkan kata dari kalimat utuh menjadi beberapa kata yang terpisah yang ada pada dokumen yang telah diinputkan di proses sebelumnya.

2.3 Cleaning Text

Cleaning text merupakan tahap ketiga dari proses pemeriksaan struktur kalimat secara otomatis. Cleaning text merupakan proses pengolahan teks dengan tujuan untuk membersihkan teks dari karakter atau elemen yang tidak dibutuhkan. Pada langkah ini, tindakan yang akan diambil adalah menghilangkan semua tanda baca yang ada dalam data. Data yang digunakan merupakan hasil dari proses pemisahan data pada tahap sebelumnya.

2.4 Melakukan POS Tagging

POS (Part Of Speech) Tagging adalah tahap keempat dalam proses pemeriksaan struktur kalimat secara otomatis. *POS (Part Of Speech) Tagging* adalah proses di mana label diberikan secara otomatis pada kata-kata dalam sebuah kalimat (Hari Wijayanto and Setio Pribadi 2022). Secara menyeluruh, label kata dapat dikelompokkan ke dalam kategori- kategori seperti Kata Benda (Noun), Kata Kerja (Verb), Kata Sifat (Adjective), Kata Keterangan (Adverb), Kata Penghubung (Conjunction Word), Kata Bilangan (Numeral), dan juga mencakup tanda baca lainnya (Aprilliwanto, Sanjaya, and Widodo 2021). Pada tahap ini, pemeriksaan dilakukan dengan menggunakan *POS (Part Of Speech) Tagging* menggunakan *Flair NLP (Natural Language Processing)*. *Flair NLP* merupakan salah satu perangkat perpustakaan NLP yang digunakan dalam analisis bahasa. *Flair NLP* telah menyediakan koleksi teks dalam bahasa Indonesia yang dapat digunakan untuk *POS (Part Of Speech) Tagging* dan setiap kata akan dicocokkan dengan perpustakaan yang digunakan. Label tersebut akan ditempatkan di sebelah kanan kata yang sedang dianalisis.

2.5 Melakukan Pemeriksaan dengan ANTLR Parser

ANTLR (ANother Tool Language Recognition) adalah sebuah alat pembuat parser yang diciptakan untuk membaca, memproses, mengeksekusi, atau menerjemahkan teks yang memiliki struktur tertentu (Natali and Alfadian 2019). Peran *ANTLR (ANother Tool Language Recognition)* adalah sebagai pengurai sintaks (parser) dalam proses mengevaluasi kode. Generator *ANTLR (ANother Tool Language Recognition) Parser* menggunakan strategi parsing top-down yang mendasarinya (Parr and Fisher 2011). Mesin akan mengikuti pola membaca dari atas ke bawah dan dari kiri ke kanan terlebih dahulu pada struktur yang sering disebut sebagai “tree walker” (Tresnayatna, Sri Widowati, and Lukmanul Hakim 2019). Walking tree memiliki peran dalam mengidentifikasi setiap tokenisasi dalam pohon penguraian, namun keduanya memiliki aturan yang berbeda-beda tergantung pada keperluan tata bahasa dan aturan serta hasil keluaran yang diinginkan. *ANTLR (ANother Tool Language Recognition)* memerlukan aturan tata bahasa (grammar) yang spesifik untuk setiap bahasa pemrograman yang digunakan. Kemampuan ini memungkinkan kita untuk menciptakan bahasa pemrograman sendiri dengan aturan dan tata bahasa yang dapat disesuaikan sesuai kebutuhan. Pada tahap keempat ini terdapat proses “Lexer” dan proses “Parser”. Langkah awalnya adalah kode program akan diurai menjadi token menggunakan lexer. Berikutnya, Parser akan menggunakan token-token tersebut untuk menciptakan pohon sintaks. Pohon sintaks adalah struktur data yang memperlihatkan hubungan antara elemen-elemen dalam kode program, seperti operator, variabel, dan fungsi (Adiyoso and Susetyo 2023). Pemanfaatan pohon sintaks mempermudah proses analisis pada kode program. Selain itu, pohon sintaks juga mendukung pembentukan kode yang terstruktur dan mudah dipahami. Gambar 2. menggambarkan proses *ANTLR (ANother Tool Language Recognition) Parser* yang dijelaskan dalam jurnal (Adiyoso and Susetyo 2023).



Gambar 2. Diagram Proses ANTLR Parser

Pemeriksaan melibatkan 4 jenis struktur kalimat, yaitu S-P, S-P-O, S-P-K, dan S-P-O-K. Setiap struktur kalimat ini terdiri dari beragam komponen pembentuk yang berbeda.

2.6 Menyimpan Hasil Pemeriksaan

Tahap ini merupakan langkah terakhir dalam pemeriksaan otomatis terhadap struktur kalimat. Pada tahap ini, hasil pemeriksaan SPOK akan disimpan dalam format file .xlsx. Proses pengelompokan kalimat yang mengalami kesalahan dan perhitungan tingkat keberhasilan metode ini akan dicatat dalam dokumen dengan format file .txt.

2.7 Hasil Pengujian

Pada langkah ini, dilakukan analisis hasil uji coba. Analisis ini akan didasarkan pada 1 uji coba, yang bertujuan untuk menghitung tingkat keberhasilan pemeriksaan struktur kalimat dalam sebuah teks.

2.7.1 Tingkat Keberhasilan

Pengujian tingkat keberhasilan merupakan evaluasi yang dilakukan untuk memperoleh persentase tingkat keberhasilan dalam memeriksa struktur kalimat dalam sebuah teks. Langkah-langkah dalam pengujian ini mencakup menghitung jumlah kalimat yang ada dan mengevaluasi keluaran berupa kalimat yang telah terstruktur. Selanjutnya, persentase keberhasilan dihitung dengan membagi total jumlah kalimat yang tersedia dengan jumlah kalimat yang telah terstruktur. Keberhasilan pemeriksaan ini dapat dinilai melalui kesesuaian kalimat-kalimat dengan struktur yang telah ditetapkan dalam analisis manual.

$$\text{Tingkat Keberhasilan} = \frac{\epsilon \text{ Jumlah Kalimat Berstruktur}}{\epsilon \text{ Jumlah Kalimat Keseluruhan}} \quad \dots(1)$$

$$\text{Keberhasilan (\%)} = \text{Tingkat Keberhasilan} \times 100\%$$

3. HASIL DAN PEMBAHASAN

Pemeriksaan struktur kalimat yang dilakukan secara otomatis ini memeriksa sebanyak 500 kalimat. Kalimat tersebut merupakan kalimat tunggal. Hasil dari penerapan *POS (Part Of Speech) Tagging* dengan *ANTLR (ANother Tool Language Recognition) Parser* dalam memeriksa struktur kalimat bahwa terdapat 424 kalimat yang benar dari 500 kalimat dengan 0,848 atau tingkat keberhasilan (%) 84,8% dan ada 76 kalimat yang tidak berhasil didefinisikan strukturnya. Kalimat yang tidak berhasil tersebut diklasifikasikan berdasarkan penyebab dari ketidakberhasilannya tersebut. Informasi tentang hasil pemeriksaan keseluruhan terhadap struktur kalimat dapat ditemukan dalam Tabel 1.

Tabel 1. Hasil Pemeriksaan SPOK Otomatis

No	Hasil Pemeriksaan Otomatis	Jumlah Frasa
1.	Berhasil	424 Kalimat
2.	Gagal, karena ada kesalahan pengelompokkan kelas kata pada bagian kata kerja. Kata kerja yang seharusnya termasuk ke dalam kelas kata “VERB” akan tetapi menurut algoritma ini termasuk ke dalam kelas kata “NOUN”	11 Kalimat
3.	Gagal, karena ada kesalahan pengelompokkan kelas kata pada bagian kata kerja. Kata kerja yang seharusnya termasuk ke dalam kelas kata “VERB” akan tetapi menurut algoritma ini termasuk ke dalam kelas kata “ADP”	1 Kalimat
4.	Gagal, karena kelas kata “SCONJ” belum didefinisikan di dalam grammar.	10 Kalimat
5.	Gagal, karena ada kesalahan pengelompokkan kelas kata pada bagian subjek. Subjek yang seharusnya termasuk ke dalam kelas kata “SCONJ” akan tetapi menurut algoritma ini termasuk ke dalam kelas kata “NOUN”.	1 Kalimat
6.	Gagal, karena kelas kata “CCONJ” belum didefinisikan di dalam grammar.	2 Kalimat
7.	Gagal, karena hasil pemeriksaan struktur kalimat yang dilakukan secara otomatis berbeda dengan hasil pemeriksaan yang dilakukan secara manual.	40 Kalimat

8.	Gagal, karena kalimat tersebut termasuk ke dalam kalimat kompleks.	1 Kalimat
9.	Gagal, karena pola NOUN-ADP tidak ada di dalam grammar.	3 Kalimat
10.	Gagal, karena pola NOUN-ADJ tidak ada di dalam grammar.	4 Kalimat
11.	Gagal, karena di dalam grammar belum mengadopsi frase dengan kata sandang di awal.	3 Kalimat

Terdapat 8 klasifikasi penyebab dari kegagalan dari proses tersebut, berikut ini penjelasannya :

Gagal, karena terdapat kesalahan pada bagian pengelompokkan kelas kata “VERB”. Kata kerja yang mana seharusnya termasuk ke dalam kelas kata “VERB”, akan tetapi hasil dari proses Post Tagging kata kerja termasuk ke dalam kelas kata “NOUN”. Contoh : kata “memasak” pada kalimat “Ayah memasak masakan lezat di dapur”. Kata “memasak” seharusnya masuk ke dalam kelas kata “VERB” karena kata tersebut merupakan kata kerja, akan tetapi hasil dari proses *POS (Part Of Speech) Tagging* kata “memasak” termasuk ke dalam kelas kata “NOUN”.

Gagal, karena terdapat kesalahan pada bagian pengelompokkan kelas kata. Kata kerja yang mana seharusnya termasuk ke dalam kelas kata “VERB”, akan tetapi hasil dari proses *POS (Part Of Speech) Tagging* kata kerja termasuk ke dalam kelas kata “ADP”. Contoh : kata “mempratikkan” pada kalimat “Penari tari tradisional mempratikkan gerakan-gerakan yang rumit”. Kata “mempratikkan” seharusnya masuk ke dalam kelas kata “VERB” karena kata tersebut merupakan kata kerja, akan tetapi hasil dari proses *POS (Part Of Speech) Tagging* kata “mempratikkan” termasuk ke dalam kelas kata “ADP”.

Gagal, karena kelas kata “SCONJ” belum didefinisikan di dalam grammar. Contoh : kata “Siswa-siswa” pada kalimat “Siswa-siswa menulis esai untuk lomba”. Kata “Siswa-siswa” termasuk dalam kelas kata “SCONJ”, tetapi kelas kata “SCONJ” belum didefinisikan di dalam grammar, maka yang terjadi adalah kalimat tersebut

tidak dapat ditentukan SPOK nya.

Gagal, karena ada kesalahan dalam pengelompokkan kelas kata. Kata yang seharusnya masuk ke dalam kelas kata “SCONJ”, namun hasil dari proses *POS (Part Of Speech) Tagging* mengkategorikannya sebagai kelas kata “NOUN”. Sebagai contoh : kata “Anak-anak” dalam kalimat “Anak-anak di desa ini sering bermain di lapangan tanah”.

Kata “Anak-anak” seharusnya merupakan kelas kata “SCONJ”, tetapi dalam hasil *POS (Part Of Speech) Tagging* kata “Anak-anak” diklasifikasikan sebagai kelas kata “NOUN”.

Gagal, karena kelas kata “CCONJ” belum didefinisikan di dalam grammar. Contoh pada kalimat “Astronot menjalani pelatihan fisik dan mental”. Kalimat tersebut mengandung kata “dan” yang mana kata tersebut termasuk ke dalam kelas kata “CCONJ”, akan tetapi kelas kata “CCONJ” belum didefinisikan di dalam grammar, maka yang terjadi adalah kalimat tersebut tidak dapat didefinisikan SPOK nya.

Gagal, karena hasil pemeriksaan struktur kalimat yang dilakukan secara otomatis berbeda dengan hasil pemeriksaan yang dilakukan secara manual. Contoh pada kalimat “Petugas kebersihan membersihkan taman kota setiap pagi”. Jika kalimat tersebut diperiksa secara manual akan menghasilkan output “Petugas kebersihan (S) membersihkan (P) taman kota (O) setiap pagi (K)”, akan tetapi pemeriksaan secara otomatis menghasilkan output “Petugas kebersihan (S) membersihkan (P) taman kota setiap pagi (O)”.

Gagal, karena kalimat tersebut merupakan kalimat kompleks. Contoh pada kalimat “Gajah itu memiliki belalai yang panjang dan kuat”. Kalimat “yang panjang dan kuat” merupakan anak kalimat dari sebuah keterangan. Sehingga, kalimat tersebut tidak dapat didefinisikan dengan baik SPOK nya karena kalimat tersebut termasuk ke dalam kalimat kompleks.

Gagal, karena di dalam grammar belum mencakup kalimat dengan pola yang sesuai dengan kalimat tersebut. Contoh pada kalimat “Pohon apel di kebun itu menghasilkan buah-buah yang manis”. Di dalam kalimat tersebut terdapat pola NOUN ADP yang mana pohon (NOUN) apel (NOUN) di (ADP) sedangkan di dalam grammar tidak ada terdapat pola yang sesuai dengan pola kalimat tersebut. Sehingga, kalimat

tersebut tidak dapat didefinisikan dengan baik SPOK nya.

Gagal, karena di dalam grammar belum mencakup kalimat dengan pola yang sesuai dengan kalimat tersebut. Contoh pada kalimat “Silvera cantik”. Di dalam kalimat tersebut terdapat pola NOUN-ADJ yang mana Silvera (NOUN) cantik (ADJ) sedangkan di dalam grammar tidak ada terdapat pola yang sesuai dengan pola kalimat tersebut. Sehingga, kalimat tersebut tidak dapat didefinisikan dengan baik SPOK nya.

Gagal, karena pada grammar belum mengadopsi frase dengan kata sandang di awal. Contoh : kata “Para” pada kalimat “Para peselancar menantang ombak besar dengan papan selancar mereka”. Kata “Para” di awal kalimat termasuk dalam kelas kata “DET” yang merupakan kata penunjuk. Oleh karena itu, algoritma ini belum dapat menentukan pola dengan akurat.

Beberapa faktor mempengaruhi hasil dari pengujian tingkat keberhasilan tersebut (Hari Wijayanto and Setio Pribadi 2022). Faktor awal adalah output yang dihasilkan oleh proses *POS (Part Of Speech) Tagging* yang memanfaatkan Flair NLP sebagai sumber data. Keluaran dari *POS (Part Of Speech) Tagging* bisa bervariasi tergantung pada posisi kata dalam kalimat. Sebagai contoh kata “memasak” dalam kalimat “Ayah memasak masakan lezat di dapur” mungkin akan diberi label jenis kata <NOUN> atau kata benda. Namun, sebenarnya, kata “memasak” termasuk dalam kategori kata <VERB> atau kata kerja. Variasi ini mungkin menyebabkan kesalahan dalam langkah-langkah berikutnya.

Salah satu faktor lainnya adalah variasi komponen tag yang digunakan untuk membentuk pola pada setiap kalimat. Dalam penelitian ini, peneliti membatasi pola kalimat yang digunakan. Peneliti hanya menggunakan 4 pola kalimat, yakni “S- P”, “S- P-O”, “S-P-K”, dan “S-P-O-K”. Pola- pola kalimat tersebut termasuk dalam kategori kalimat tunggal. Saat ini, tidak ada pola kalimat yang tersedia untuk memeriksa jenis kalimat majemuk dan jenis kalimat lainnya. Dengan kata lain, ketika memeriksa kalimat selain kalimat tunggal, kemungkinan akan muncul kesalahan atau kalimat tersebut tidak dapat didefinisikan ke dalam pola atau struktur kalimat, meskipun pemeriksaan manual menghasilkan pola atau struktur kalimat yang benar.

Dalam penelitian yang berjudul “Pemeriksaan Pola Kalimat Otomatis Pada Sebuah Karangan Menggunakan *POS (Part Of Speech) Tagging* Bahasa Indonesia Dan LALR Parser” dinyatakan bahwa penerapan *POS (Part Of Speech) Tagging*

dengan LALR Parser untuk memeriksa pola kalimat secara otomatis menghasilkan tingkat keberhasilan sebesar 79,2%. Dalam penelitian ini, peneliti menguji struktur kalimat secara otomatis menggunakan *POS (Part Of Speech) Tagging* dengan *ANTLR (ANother Tool Language Recognition) Parser*. Hasilnya menunjukkan bahwa struktur kalimat dapat diperiksa dengan benar untuk 424 kalimat dari total 500 kalimat yang dimasukkan. Temuan ini menunjukkan bahwa penelitian ini mencatat tingkat keberhasilan rata-rata sebesar 0,848 atau 84,8% dalam hal keberhasilan secara total.

4. PENUTUP

Penggunaan metode *POS (Part Of Speech) Tagging* dengan *ANTLR (ANother Tool Language Recognition) Parser* dapat mendukung pemeriksaan otomatis terhadap struktur kalimat dalam sebuah kalimat. Dalam penelitian ini, ditemukan bahwa metode tersebut berhasil memeriksa 424 kalimat dengan benar dari total 500 kalimat yang diuji, dengan tingkat keberhasilan sebesar 0,848 atau persentase keberhasilan sebesar 84,8%. Beberapa faktor mempengaruhi hasil penelitian ini, termasuk penggunaan Flair NLP untuk memberi label jenis kata dan batasan penelitian hanya untuk penggunaan kalimat tunggal. Rekomendasi untuk penelitian berikutnya adalah mempertimbangkan penggunaan perpustakaan NLP atau korpus yang lebih komprehensif dan akurat dalam menetapkan label jenis kata daripada perpustakaan Flair NLP yang digunakan dalam penelitian ini, serta mempertimbangkan perluasan struktur kalimat yang digunakan. Penggunaan tersebut tidak hanya terbatas pada kalimat tunggal, tetapi juga mencakup jenis kalimat lain seperti kalimat majemuk dan jenis kalimat lainnya, sehingga struktur kalimat dapat diperiksa dan diidentifikasi dengan lebih baik.

DAFTAR PUSTAKA

- Abdiansah, and Muhammad Qurhanul Rizqie. 2023. "Automatic Language Identification For Indonesia-Malaysian Language Using Machine Learning." *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika* 9(2): 104–10.
- Adiyoso, Wibisono, and Yeremia Alfa Susetyo. 2023. "Pembangunan Compiler Domain Specific Language Sebagai Generator Form Html Menggunakan Python Sly." *Jurnal Indonesia : Manajemen Informatika dan Komunikasi* 4(2): 449–61.
- Ankali, Sanjay B., and Latha Parthiban. 2021. "Detection and Classification of Cross-

- Language Code Clone Types by Filtering the Nodes of ANTLR-Generated Parse Tree.” *International Journal of Intelligent Systems and Applications* 13(3): 43–65.
- Aprilliwanto, Rino Ekta, Ardi Sanjaya, and Danang Wahyu Widodo. 2021. “Identifikasi Pola Kalimat Bahasa Indonesia Pada Siswa Sekolah Dasar Menggunakan Metode LALR Dan Stemming.” *SEMNAS Inovasi Teknologi UN PGRI Kediri*: 119–26.
- Dan, Komputer, and Sistem Intelijen. 2012. “Teknologi Informasi Dan Komunikasi (Tik) Untuk Ketahanan Nasional.” *Komputer Dan Sistem Intelijen* 7(2302–3740): 311–16. <http://penelitian.gunadarma.ac.id/kommit>.
- Fauzia, Sitti. 2018. “BERBICARA SEBAGAI SUATU KETERAMPILAN BERBAHASA.” : 282.
- Hari Wijayanto, Sukmandaru, and Feddy Setio Pribadi. 2022. “Pemeriksaan Pola Kalimat Otomatis Pada Sebuah Karangan Menggunakan POS Tagging Bahasa Indonesia Dan LALR Parser.” *Jtim* 2022 4(3): 149–67. <https://doi.org/10.35746/jtim.v4i3.263>.
- Haryanti, Dwi et al. 2020. “Translation Method of English Into Indonesian Used in Movie Scripts and Their Application in Translation Teaching.” 461(Icllae 2019): 294–98.
- Izzati, Gita Nur. 2016. “Analisis Kesalahan Pola Kalimat Bahasa Indonesia Pada Karangan Narasi Siswa Kelas Iv Sd Negeri Bandasari Kabupaten Tegal.”
- Ludiani, Daszharah, M.Pd. Dr. Laili Rahmawati, S.Pd., and M.Hum Dr. Atiqa Sabardila. 2022. “Pengembangan Media Pembelajaran Interaktif Dalam Pembelajaran Keterampilan Berbahasa Indonesia Dengan Pendekatan Whole Language.” *Universitas Muhammadiyah Surakarta*: 1–20.
- Martha, Nila, and Yehonala Situmorang. 2018. “Meningkatkan Kemampuan Menulis Siswa Melalui Teknik Guiding Questions.” *Journal of Education Action Research* 2(2): 165–71. <https://ejournal.undiksha.ac.id/index.php/JEAR/index>.
- Mizkat, Eva. 2019. “Analisis Penggunaan Kalimat Efektif Pada Penulisan Kritik Dan Saran Mahasiswa FKIP UNA.” *Jurnal Komunitas Bahasa* 7(1): 19–32. <http://jurnal.una.ac.id/index.php/jkb/article/view/784/0>.
- Natali, Vania, and Pascal Alfadian. 2019. “Analisis Dan Perancangan Domain Specific Language Untuk Data Generator Pada Relational Database.” *JUMANJI (Jurnal Masyarakat Informatika Unjani)* 3(01): 64.
- Ortin, Francisco, Jose Quiroga, Oscar Rodriguez-Prieto, and Miguel Garcia. 2022. “An Empirical Evaluation of Lex/Yacc and ANTLR Parser Generation Tools.” *PLoS ONE* 17(3 March): 1–16. <http://dx.doi.org/10.1371/journal.pone.0264326>.
- Parr, Terence, and Kathleen Fisher. 2011. “LL(*): The Foundation of the ANTLR Parser Generator.” *Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*: 425–36.

- Pohan, Desi Damayani, and Ulfi Sayyidatul Fitria. 2021. "Jenis Jenis Komunikasi." *Journal Educational Research and Social Studies* 2: hal. 31.
- Pratama, Mudafiq Riyan, Sri Kusumadewi, and Taufiq Hidayat. 2017. "Penerapan Algoritma Lalr Parser Dan Context-Free Grammar Untuk Struktur Kalimat Bahasa Indonesia." *Jurnal Teknologi Elektro* 8(1): 1.
- Ramadhanti, Febyana, Yudi Wibisono, and Rosa Ariani Sukamto. 2019. "Analisis Morfologi Untuk Menangani Out-of-Vocabulary Words Pada Part-of-Speech Tagger Bahasa Indonesia Menggunakan Hidden Markov Model." *Jurnal Linguistik Komputasional (JLK)* 2(1): 6.
- Rusdianto, Stephanie, and Ria Chaniago. 2018. "Penerapan Abstract Syntax Tree Dan Algoritme Damerau-Levenshtein Distance Untuk Mendeteksi Plagiarisme Pada Berkas Source Code." *Jurnal Telematika* 13 13(2): 105–10.
- Sa, Dian, Yogan Jaya Kumar, and Fauziah Binti Kasmin. 2023. "Combination of Graph-Based Approach and Sequential Pattern Mining for Extractive Text Summarization with Indonesian Language." 9(2): 132–45.
- Tresnayatna, Bayusandya, Ir Sri Widowati, and Iman Lukmanul Hakim. 2019. "Pembangkit Test Case Untuk Pengujian Perangkat Lunak Menggunakan Metode Basis Path." *e-Proceeding of Engineering* 6(1): 2189.
- Wijanarto, and Ajib Susanto. 2014. "Aspek Pedagogik Implementasi Translator Notasi Algoritmik Berbasis Parsing Ll (*) Dan String." *Techno.COM* 13(1): 1–8.
- Yazid, Ahmad Subhan, and Agung Fatwanto. 2018. "Penentuan Kelas Kata Pada Part of Speech Tagging Kata Ambigu Bahasa Indonesia." *JISKA (Jurnal Informatika Sunan Kalijaga)* 2(3): 157.