

NATURAL LANGUAGE PROCESSING UNTUK OTOMATISASI PENGENALAN PRONOMINA DALAM KALIMAT BAHASA INDONESIA

Mohammad Farid Naufal¹, Selvia Ferdiana Kusuma*²

¹Universitas Surabaya, Surabaya, ²Politeknik Elektronika Negeri Surabaya, Surabaya
Email: ¹faridnaufal@staff.ubaya.ac.id, ²*selvia@pens.ac.id,
*Penulis Korespondensi

(Naskah masuk: 29 Juni 2022, diterima untuk diterbitkan: 24 Oktober 2022)

Abstrak

Pronomina (kata ganti) adalah jenis kata yang dapat dipakai untuk menggantikan posisi kata benda atau orang dalam suatu kalimat. Penggunaan pronomina akan mudah dipahami apabila serangkaian kalimat dibaca secara utuh. Namun jika rangkaian kalimat tersebut hanya dibaca pada kalimat-kalimat tertentu, maka akan sulit memahami kalimat yang memiliki pronomina. Pada pengolahan bahasa alamiah, diperlukan kejelasan konteks dari sebuah kalimat. Dalam konteks otomatisasi pengolahan bahasa alamiah, adanya pronomina dapat menyulitkan komputer untuk memahami kalimat tersebut. Oleh sebab itu, dalam pengolahan bahasa alamiah yang mengandung pronomina diperlukan pre proses berupa pengubahan pronomina ke dalam bentuk subjek atau objek asli yang dirujuk. Metode yang diusulkan untuk menyelesaikan permasalahan ini adalah pendekatan berbasis sintaktik. Pendekatan ini menitikberatkan pada struktur kata yang digunakan dan struktur komponen kata yang digunakan. Metode yang diusulkan memiliki 4 tahapan yakni pengumpulan data, pembangkitan aturan, otomatisasi pengenalan pronominal, dan terakhir adalah evaluasi. Metode yang diusulkan telah diujicobakan untuk mengenali adanya pronomina dari kalimat-kalimat pada materi Ilmu Pengetahuan Alam dan Ilmu Pengetahuan Sosial di jenjang sekolah dasar. Hasil evaluasi menunjukkan bahwa metode yang diusulkan dapat digunakan untuk mengubah subjek yang berbentuk pronomina menjadi subjek atau objek asli yang dirujuk. Rata-rata akurasi yang didapatkan sebesar 81%. Akurasi tersebut didapatkan dari perbandingan antara jumlah kata ganti yang berhasil diidentifikasi subjeknya dengan keseluruhan data uji. Hasil dari penelitian ini dapat digunakan peneliti di bidang *Natural Language Processing* untuk melakukan pra proses terhadap teks yang akan diolah.

Kata kunci: bahasa Indonesia, otomatisasi, pengenalan pronomina.

NATURAL LANGUAGE PROCESSING FOR AUTOMATION OF PRONOMINAL RECOGNITION IN INDONESIAN SENTENCES

Abstract

A pronoun is a word that can be used to replace a noun or person in a sentence. The use of pronouns will be easy to understand if a series of sentences is read in its entirety. However, if the sentence series is only read in specific sentences, it will be difficult to understand sentences with pronouns. In natural language processing, it is necessary to clarify the context of a sentence. In the context of automation of natural language processing, the existence of pronouns can make it difficult for computers to understand the sentence. Therefore, in processing natural language containing pronouns, it is necessary to pre-process in the form of converting pronouns into the form of the original subject or object referred to. The method proposed to solve this problem is a syntactic-based approach. This approach focuses on the structure of the words used and the word components used. The proposed method has 4 stages, namely data collection, rule generation, automation of pronoun recognition and the last is evaluation. The proposed method has been evaluated to identify the existence of pronouns from sentences in the Natural Sciences and Social Sciences material at the elementary school level. The evaluation results show that the proposed method can be used to change the subject in the form of a pronoun into the original subject or object referred to. The average accuracy obtained is 81%. The accuracy is obtained from the comparison between the number of pronouns that have been identified with the overall test data. Researchers in natural language processing can use the results of this study to pre-process their text.

Keywords: automation, Indonesian, pronoun recognition

1. LATAR BELAKANG

Saat ini, penelitian tentang *Natural Language Processing* (NLP) semakin berkembang. Berbagai macam penelitian tersebut menggunakan teks sebagai datanya. Misalnya, penelitian tentang *chatbot* (Khoirunisa *et al.*, 2020), analisis sentimen (Kastrati *et al.*, 2021), peringkasan teks/dokumen (Awasthi *et al.*, 2021), pembangkitan pertanyaan (Pistol, Trandabät and Räschip, 2018) dan pembangkitan ontologi informasi (Kusuma, Sukya and Heriadi, 2021). Semua kajian tersebut fokus pada pengolahan teks untuk menghasilkan berbagai hal sesuai dengan tujuannya, misalnya dalam penelitian tentang otomatisasi pembentukan ontologi. Komputer harus dapat memahami semua kalimat yang dijadikan masukan untuk dapat merelasikan semua informasi. Namun, tidak semua kalimat dapat dipahami dan direlasikan dengan mudah. Komputer akan lebih sulit memahami kalimat yang memiliki pronomina daripada kalimat yang tidak memiliki pronomina. Pronomina atau yang biasa dikenal dengan “kata ganti” dalam sebuah kalimat, adalah kata yang berfungsi sebagai pengganti kata benda (Moeliono *et al.*, 2017). Contoh kalimat yang memiliki pronomina adalah “Mereka akan pergi ke Surabaya siang ini”. Kalimat tersebut tidak dapat dikaitkan dengan subjek atau objek apa pun karena subjek atau objek sebenarnya yang dirujuk tidak diketahui. Pronomina “mereka” harus diubah menjadi subjek atau objek asli yang dirujuknya.

Penelitian dari Aloraini dan Chang (Chang *et al.*, 2017; Aloraini and Poesio, 2020) juga menjelaskan bahwa pentingnya *zero pronoun* dalam *Natural Language Processing*. Kedua penelitian tersebut menggunakan metode *deep learning* untuk melakukan pengenalan pronominal pada bahasa Cina dan Arab. Metode tersebut dipilih karena memang telah ada dataset yang cukup memadai yang dapat digunakan untuk dijadikan data latih. Namun dataset serupa belum ada untuk bahasa Indonesia. Di Indonesia, ada dua penelitian membahas tentang pengenalan pronomina yang kemudian digunakan untuk mendukung pembangkitan pertanyaan. Penelitian dari Saripudin (Saripudin and Purnamasari, 2017) menjelaskan pengenalan pronomina dalam sebuah kalimat untuk mengurangi ambiguitas pertanyaan. Metode yang digunakan adalah pengenalan subjek pada kalimat sebelumnya. Hasil dari proses ini tidak dijelaskan secara rinci karena fokusnya adalah pada proses pembangkitan pertanyaan. Penelitian kedua dilakukan oleh Kusuma (Kusuma and Alhamri, 2018). Penelitian ini juga melakukan proses pengenalan pronomina untuk menghasilkan pertanyaan yang tidak ambigu. Pronomina yang dikenali dalam penelitian itu hanya “ini” dan “itu”. Metode yang digunakan adalah dengan mengenali kata sebelum kata “adalah” atau “merupakan” dari sebuah kalimat. Kata yang muncul sebelum dua kata kunci tersebut dianggap sebagai subjek asli dari pronomina yang ada di kalimat

selanjutnya. Hasil evaluasi tentang pengenalan pronomina dalam penelitian ini tidak dibahas secara rinci. Namun, penelitian ini menjelaskan bahwa pengenalan pronomina dapat meningkatkan akurasi dan kualitas hasil pertanyaan. Kedua penelitian tersebut hanya menjelaskan secara singkat pengenalan pronomina menggunakan subjek dari kalimat sebelumnya dan struktur kata yang digunakan pada kalimat sebelumnya. Padahal, jenis dan letak pronomina yang digunakan dalam setiap kalimat bisa berbeda.

Oleh karena itu, penelitian ini membahas tentang otomatisasi pengenalan pronomina yang lebih beragam, dengan harapan dapat membantu pre-proses data pada penelitian selanjutnya di bidang *Natural Language Processing*.

2. METODE PENELITIAN

Ada empat tahapan dalam penelitian ini, yaitu pengumpulan dan analisis data, penentuan aturan, otomatisasi pengenalan pronomina, dan yang terakhir adalah evaluasi. Detail dari masing-masing proses dijelaskan pada sub bab 2.1 sampai dengan sub bab 2.4.

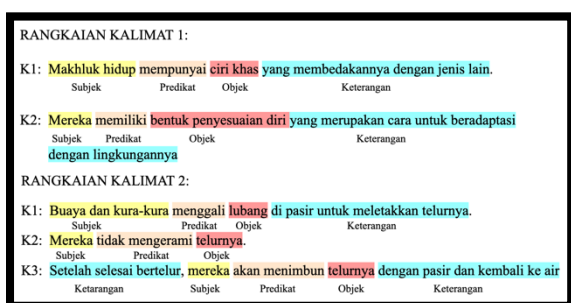
2.1. Pengumpulan dan Analisis Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah *purposive sampling*. *Purposive sampling* adalah metode penentuan sampel dengan pertimbangan tertentu [9]. Metode pengumpulan data ini dipilih karena struktur kalimat bahasa Indonesia yang mengandung pronomina sangat beragam dan tidak dapat terpola dengan pasti. Oleh karena itu, pemilihan data akan dilakukan dengan pertimbangan tertentu agar data yang digunakan dapat mewakili sebagian besar tipe data yang sebenarnya. Data yang digunakan dalam penelitian ini adalah materi pelajaran Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS) kelas 4, 5, dan 6 Sekolah Dasar. Data yang digunakan adalah rangkaian kalimat dengan pronomina, contohnya “Makhluk hidup memiliki ciri khas yang membedakannya dengan jenis lain. Mereka memiliki penyesuaian diri yang merupakan cara untuk menyesuaikan bentuk dengan lingkungannya”

Semua kalimat selanjutnya dianalisis berdasarkan jenis pronomina, subjek atau objek yang dirujuk dari pronomina, dan struktur kalimat pronomina. Analisis dilakukan oleh guru di bidang Ilmu Pengetahuan Alam, Ilmu pengetahuan Sosial dan juga dosen bahasa Indonesia yang memiliki pengalaman mengajar lebih dari lima tahun. Rangkaian kalimat yang telah dianalisis kemudian digunakan sebagai dasar untuk menentukan kaidah pengenalan pronomina. Penjelasan tentang proses analisis dijelaskan lebih detail pada sub bab pembangkitan aturan.

2.2. Pembangkitan Aturan

Pengenalan pronomina dilakukan berdasarkan aturan. Pertama, aturan dibentuk berdasarkan hasil analisis kalimat yang telah dilakukan. Cara membentuk aturannya adalah dengan mengelompokkan beberapa kalimat dengan pronomina yang sama kemudian menentukan subjek atau objek asli yang dimaksud. Setelah itu dilakukan analisis hubungan antar jenis pronomina, letak struktur pronomina atau subjek yang dimaksud dalam rangkaian kalimat tersebut. Ilustrasi proses analisis di tunjukkan pada Gambar 1.



Gambar 1. Ilustrasi Analisis Kalimat

Warna kuning mewakili subjek, warna oranye mewakili predikat, warna merah mewakili objek, dan warna biru muda mewakili informasi dalam sebuah kalimat. Rangkaian kalimat pertama adalah “Makhluk hidup memiliki ciri khas yang membedakannya dengan jenis lain. Mereka memiliki penyesuaian diri yang merupakan cara untuk menyesuaikan bentuk dengan lingkungannya”. Rangkaian kalimat kedua adalah “Buaya dan kura-kura menggali lubang di pasir untuk meletakkan telurnya. Mereka tidak mengerami telurnya. Setelah selesai bertelur, mereka akan menimbun pasir itu dan kembali ke air”. Berdasarkan dua contoh rangkaian kalimat pada Gambar 1, diketahui bahwa kedua rangkaian kalimat tersebut mengandung pronomina “mereka”. Pronomina "mereka" pada rangkaian kalimat 1 mengacu pada subjek pada kalimat pertama, yaitu "makhluk hidup". Sedangkan pronomina “mereka” pada rangkaian kalimat 2 mengacu pada subjek kalimat pertama, yaitu “buaya dan kura-kura”. Dari kedua contoh tersebut terlihat bahwa tidak semua subjek pronomina ditemukan langsung pada kalimat sebelumnya. Namun dapat ditemukan pada kalimat pertama dari rangkaian kalimat. Selain itu, ada juga pola yang menunjukkan bahwa pronomina “mereka” selalu mengacu pada subjek dalam kalimat sebelum pronomina ditemukan. Dari analisis ini dapat ditemukan pola bahwa jika pronomina “mereka” ditemukan, maka diambil subjek yang tidak berbentuk pronomina pada kalimat sebelumnya untuk menggantikan pronomina “mereka”. Metode analisis aturan seperti ini juga dapat digunakan untuk menghasilkan aturan dari pronomina lainnya.

2.3. Pengenalan Pronomina

Pronomina adalah kata yang berfungsi sebagai pengganti kata benda (Moeliono *et al.*, 2017). Ada enam jenis pronomina: pesona penyapa, penunjuk, penanya, tak tentu, dan jumlah (Moeliono *et al.*, 2017). Detail dari jenis, deskripsi, dan contoh masing-masing pronomina dapat dilihat pada Tabel 1.

Pronomina persona dapat digunakan untuk merujuk pada objek seseorang. Misalnya, “saya”, “aku”, “ia”, “dia”, “beliau”, “-nya”. Pronomina penyapa dapat digunakan untuk menyebut pronomina orang kedua atau sebutan di suatu daerah, lingkungan, atau budaya. Misalnya, “Pak”, “Bu”, “Prof”, “Dok”. Pronomina penunjuk dapat digunakan untuk merujuk pada penunjuk. Pronomina penunjuk dibagi menjadi 3, yaitu pronomina umum, seperti “ini” dan “itu”. Pronomina yang menunjukkan tempat, misalnya “di sana”, “di sini”, atau “di situ”. Kemudian pronomina tentang sesuatu, misalnya “begini” atau “begitu”. Pronomina penanya dapat digunakan untuk menandai pertanyaan. Misalnya, “siapa”, “apa”, “mana”, atau “mengapa”. Pronomina tak tentu dapat digunakan untuk menyebut sesuatu yang identitasnya belum jelas, misalnya “seseorang”, “sebuah” atau “sesuatu”. Sedangkan pronomina jumlah dapat digunakan untuk menyebut suatu bilangan, misalnya “semua”, “beberapa”, atau “masing-masing”.

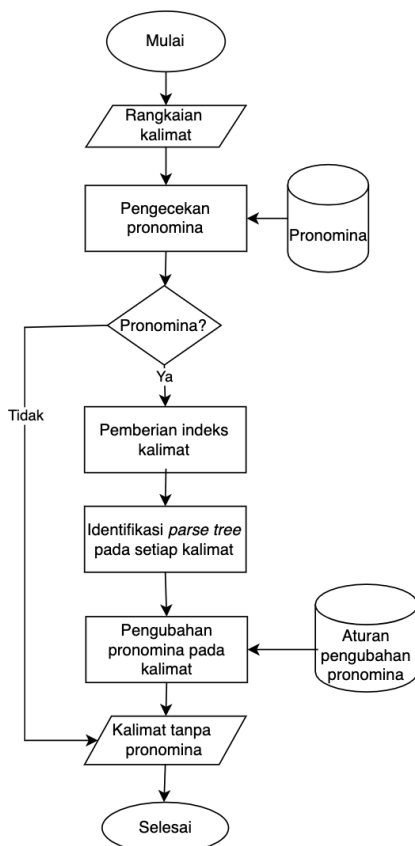
Tabel 1. Jenis Pronomina Bahasa Indonesia

JENIS	DESKRIPSI	CONTOH
Pesona	Pronomina yang mengacu pada orang	Saya, aku, ia, dia, beliau, -nya
Penyapa	Pronomina mengacu pada pronomina	Pak, Bu, Prof, Dok
Penunjuk	Pronomina yang mengacu pada lokasi atau benda.	Ini, itu, dari sini, sini, sana, begitu
Penanya	Pronomina yang digunakan sebagai penanda pertanyaan	Siapa, apa, mana, mengapa, kenapa, kapan, dimana, bagaimana, berapa
Tak tentu	Pronomina menyatakan sesuatu atau seseorang	Seorang, seekor, sesuatu, sebuah
Jumlah	Pronomina yang menyatakan jumlah	Semua, berapa, masing-masing

Pengenalan pronomina dilakukan sesuai dengan langkah-langkah yang di tunjukkan pada Gambar 2. *Input* dalam proses ini adalah rangkaian kalimat. Kemudian dilakukan pemeriksaan keberadaan pronomina. Jika tidak ditemukan, rangkaian kalimat akan langsung ditampilkan sebagai kalimat tanpa pronomina. Namun jika ditemukan pronomina maka akan masuk ke proses pemberian indeks kalimat. Pemberian indeks pada setiap kalimat bertujuan untuk memudahkan pencarian subjek pronomina. Setelah proses pengindeksan selesai, masuk ke proses identifikasi *parse tree* untuk setiap kalimat.

Parse tree adalah proses mengenali subjek (S), predikat (P), objek (O), pelengkap (Pel), dan keterangan (K) sebuah kalimat. Metode yang telah digunakan untuk membuat *parse tree* dalam bahasa

Indonesia antara lain Context-Free Grammar (CFG) (Anitia, Munarko and Azhar, 2020), Leaf Corner Parsing (Wardana, Swanita and Yohanes, 2019), dan gabungan Combined Lalr Parser dan Context-Free Grammar (Pratama, Kusumadewi and Hidayat, 2017). Penelitian ini menggunakan metode CFG untuk *parse tree*. Metode ini dipilih karena dianggap cocok untuk struktur kalimat dalam kumpulan data yang sangat beragam. Proses identifikasi *parse tree* dilakukan dengan library *NLTK* dari python. Hasil dari proses *parse tree* tersebut kemudian digunakan sebagai dasar untuk mengubah pronomina menjadi subjek atau objek asli yang dirujuk.



Gambar 2. Tahapan mengubah pronominal menjadi subjek yang dirujuk

Proses pengenalan pronomina dilakukan berdasarkan aturan yang telah dibuat dan pseudocode khusus yang telah disesuaikan dengan aturan tersebut. Contoh pseudocode yang digunakan untuk mengenali rangkaian kalimat yang memiliki pronomina "mereka" ditunjukkan pada Pseudocode 1. *Input* yang digunakan adalah beberapa kalimat yang dimasukkan ke dalam *Array*. Kemudian ada proses pencarian pronomina di setiap kalimat. Jika pronomina "mereka" ditemukan dalam sebuah kalimat, maka kalimat tersebut akan ditandai. Menurut aturan yang telah ditentukan, pronomina "mereka" dapat dikenali dari subjek kalimat sebelumnya. Oleh karena itu, proses selanjutnya adalah mencari subjek dari kalimat sebelumnya. Kemudian hasil pencarian tersebut disimpan dalam variabel "subjekasli". Kata "mereka"

dalam kalimat yang mengandung pronomina diubah berdasarkan nilai "subjekasli". Hasil dari proses ini adalah kalimat tanpa pronomina.

Pseudocode 1:

```

Input kalimat[];
Proses
Find pronomina
If (pronomina == "mereka")
    indeks = kalimat dengan pronomina
    subjekasli = subjek dari kalimat [indeks-1]
    kalimat [indeks] = ganti kata "mereka" dari kalimat [indeks] dengan subjekasli.
Hasil: kalimat [indeks]
  
```

2.4. Evaluasi

Metode yang diusulkan kemudian dievaluasi dengan cara melakukan pengenalan pronominal secara otomatis pada 100 kalimat yang terbagi dalam 50 kalimat dari materi IPA dan 50 kalimat dari materi IPS. Hasil pengenalan pronomina tersebut kemudian divalidasi oleh para ahli sesuai dengan kaidah pengenalan pronomina. Hasil evaluasi ini menghasilkan nilai akurasi proses pengenalan pronomina. Akurasi dihitung dengan menggunakan persamaan 1. *TN* adalah pronomina yang benar dikenali dan *s* adalah jumlah kalimat. Model evaluasi ini juga pernah dilakukan pada penelitian Kusuma (Kusuma *et al.*, 2018) dan Basuki (Basuki and Kusuma, 2018).

$$Akurasi = \frac{TN}{s} \times 100 \tag{1}$$

3. HASIL DAN ANALISIS

Berdasarkan analisis dari dataset yang digunakan, terdapat berbagai jenis pronomina dalam kalimat-kalimat yang terdapat pada materi IPA maupun IPS. Pronomina yang digunakan tidak terbatas pada jenis "ini" dan "itu". Selain pronomina, penelitian ini juga menambahkan kata "tersebut" pada aturan pengenalan kata. Menurut kaidah kamus bahasa Indonesia, kata "tersebut" tidak digolongkan sebagai pronomina (Sugono, 2008). Namun berdasarkan analisis, kata "tersebut" selalu digunakan untuk menyebut subjek atau objek dari kalimat sebelumnya. Oleh karena itu, kata "tersebut" juga termasuk dalam aturan pengenalan kata meskipun tidak tergolong pronomina. Proses penetapan aturan menghasilkan aturan pengenalan pronomina berdasarkan jenisnya. Aturan yang dihasilkan ditunjukkan pada Tabel 2.

Aturan dibuat berdasarkan jenis pronomina yang digunakan. Terdapat enam kaidah yang ditemukan dalam kalimat dari materi IPA dan IPS. Contoh aturan yang dibentuk untuk mengenal pronomina "mereka" adalah dengan mengambil subjek dari kalimat sebelumnya. Jika ada pronomina "mereka" maka kenali dengan subjek asli yang berasal dari subjek pada kalimat sebelumnya. Aturan yang sama juga dapat digunakan untuk pronomina "dia, ia, kita,

beliau”. Selain itu, ada beberapa aturan yang dapat digunakan beberapa pronomina sekaligus.

Tabel 2. Aturan Konversi Pronomina

	PRONO MINA	ATURAN
1	Mereka, -nya, dia, ia, beliau	Ambil subjek dari kalimat sebelumnya
2	Tersebut / ... tersebut	Ambil SPOPeK kalimat sebelumnya: jika di awal kalimat, kemudian hilangkan kata sebelum kata tersebut / Ambil subjek: jika ada kata yang sama di objek kalimat sebelumnya / Ambil kata(+1) : jika ada kata yang sama pada kalimat sebelumnya.
3	itu/ itulah / ... itu	Ambil SPOPeK Kalimat sebelumnya / kata yang sama di kalimat sebelumnya
4	Ini	Ambil objek: jika di awal kalimat setelah subjek (ambil semua kata sebelum ini) / Ambil subjek: jika ada kata yang sama di objek kalimat sebelumnya atau (akhir kalimat) / Ambil kata(+1) : jika ada kata yang sama pada kalimat sebelumnya.
5	... ini	Ambil SPOPeK kalimat sebelumnya
6	-Nya	Ambil subjek berupa NNP dari kalimat yang sama, jika tidak ada ambil kalimat sebelumnya.

Struktur kalimat yang mengandung pronomina cukup beragam, khususnya untuk pronomina "ini, itu, tersebut". Oleh karena itu, pola yang dihasilkan juga lebih dari satu. Misalnya, pronomina "ini" dapat merujuk dari subjek, objek, atau seluruh komponen SPOPeK dari kalimat sebelumnya. Bisa juga diambil dari kata setelah kata yang sama sebelum kata "ini" pada kalimat sebelumnya. Ada banyak kemungkinan tempat di mana subjek dirujuk jika pronomina adalah "ini". Oleh karena itu, perlu diketahui letak/susunan kata dalam kalimat dengan ketiga jenis pronomina tersebut. Sebagian besar aturan pengenalan pronomina dibentuk berdasarkan struktur subjek (S) predikat (P), objek (O), pelengkap (Pel), dan kata keterangan (K) sebuah kalimat. Oleh karena itu, proses *parse tree* juga berperan penting dalam pengenalan pronomina.

Evaluasi hasil pengenalan pronomina dilakukan dengan menggunakan 100 rangkaian kalimat yang berasal dari materi IPA dan IPS. Rangkaian kalimat diambil secara acak sehingga jenis pronomina yang digunakan juga beragam. Rincian hasil pengenalan pronomina dapat dilihat pada Tabel 3 dan Tabel 4. Tabel 3 merupakan hasil evaluasi pengenalan pronomina pada materi IPA. Tabel 4 merupakan hasil evaluasi pengenalan pronomina dari materi IPS.

Pada Tabel 3 dan 4, beberapa pronomina dikelompokkan menurut aturan yang digunakan saat mengenali pronomina. Jika dibandingkan antara Tabel 3 dan Tabel 4, maka dominasi pronomina pada masing-masing materi tidak sama. Misalnya untuk IPS, pronomina yang lebih dominan adalah "mereka, dia, ia, beliau". Hal tersebut mungkin terjadi karena bab yang diambil dari materi IPS ini fokus menjelaskan suatu subjek berupa orang dan suatu

peristiwa. Berbeda dengan cara penyampaian materi di IPA. Pronomina yang paling dominan adalah "ini".

Tabel 3. Hasil Evaluasi Materi IPA

PRONOMINA	JUMLAH PRONOMINA	SALAH	AKURASI (%)
Mereka, dia, ia, beliau	8	0	100 %
Tersebut / ... tersebut	9	3	67%
itu/ itulah / ... itu	7	2	71%
Ini / ... ini	16	3	81 %
-nya	10	2	80 %
TOTAL	50	9	80%

Tabel 4. Hasil Evaluasi Materi IPS

PRONOMINA	JUMLAH PRONOMINA	SALAH	AKURASI (%)
Mereka, dia, ia, beliau	22	3	86%
Tersebut / ... tersebut	7	2	71%
itu/ itulah / ... itu	6	1	83%
Ini / ... ini	13	3	77%
-nya	2	0	100%
TOTAL	50	9	82%

Rata-rata akurasi terbaik dari kedua data tersebut adalah pronomina "Mereka, dia, ia, beliau". Sedangkan rata-rata akurasi yang paling buruk berasal dari kata "tersebut". Berdasarkan analisis yang telah dilakukan, kata "tersebut" paling sulit dikenali karena terdapat kalimat yang sulit untuk dipolakan. Misalnya, "Belanda mengirim pasukan untuk menyerang kedudukan Diponegoro di Tegalrejo. Kampung tersebut dibakar dan direbut oleh Belanda". Tidak ada kata yang sama dengan objek pada kalimat sebelumnya, sehingga subjek pada kalimat sebelumnya tidak dapat digunakan. Kemudian kata-kata yang ada sebelum kata "tersebut" juga tidak memiliki kata yang sama dengan kalimat sebelumnya, maka objek kalimat sebelumnya tidak dapat digunakan untuk mengenali kata tersebut. Kemudian pola terakhir adalah mengambil SPOPeK dari kalimat sebelumnya, aturan ini tidak dapat digunakan karena tidak sesuai konteks. Contoh lain adalah dalam kalimat berikut "Diponegoro dan pasukannya telah membuat benteng pertahanan baru. Tempat tersebut tidak jauh dari kota Yogyakarta, yakni di Goa Selarong". Namun, ada juga rangkaian kalimat dengan kata "tersebut" yang bisa dikenali. Misalnya, "Bangsa Portugis datang ke Indonesia pada tahun 1521. Pada tahun tersebut, kerajaan yang kuat di Maluku adalah Ternate dan Tidore". Aturan yang dapat digunakan adalah mengambil kata setelah "tahun" pada kalimat sebelumnya. Pola ini juga dapat digunakan dalam kalimat berikut "Bangsa Eropa sangat membutuhkan rempah-rempah seperti cabe, mrica, kakao, cengkih. Rempah-rempah tersebut sangat membantu memanaskan suhu tubuh mereka" Kata "tersebut" dapat dikenali sebagai "cabe, mrica, kakao, cengkih".

Jika dibandingkan dengan penelitian terdahulu yang membahas pengenalan kata ganti (Saripudin and Purnamasari, 2017; Kusuma and Alhamri, 2018), penelitian ini lebih unggul karena dapat mengenali lebih banyak kata ganti jika dibandingkan penelitian terdahulu yang hanya mengenali pronomina jenis “ini” dan “itu”. Namun walaupun demikian, penelitian ini juga belum benar-benar bisa mengenali pronomina secara keseluruhan. Masih ada beberapa kesalahan ketika melakukan pengenalan pronomina. Contoh kesalahan pengenalan pronomina dan penyebabnya ditunjukkan pada Tabel 5.

Tabel 5. Contoh kesalahan pengenalan pronomina

PRONOMINA	KALIMAT	ALASAN
Mereka	Di daerah pertambangan Kalimantan dan Sumatera juga banyak tenaga kerja Indonesia. Kondisinya jauh lebih memprihatinkan. Mereka dipaksa bekerja di daerah pertambangan yang terpencil.	Kalimat terlalu kompleks
Tersebut	Apalagi Belanda sering campur tangan dalam urusan kerajaan di Maluku. Keadaan tersebut mendorong Thomas Matulesi (Pattimura) memimpin perlawanan.	Ada kata “apalagi” di awal kalimat sehingga membuat penggabungan kalimat menjadi kurang sesuai
Ini	Penduduk yang tidak mau bekerja dianggap pemberontak dan dihukum. Kegiatan ini disebut kerja paksa atau kerja rodi.	Konteks “kegiatan” yang dimaksudkan bersifat implisit, tidak tersirat secara langsung dari kalimat sebelumnya
Nya	Mamalia mempunyai kelenjar susu. Kelenjar susu digunakan untuk menyusui anak yang dilahirkannya	Harusnya pronominal -nya merujuk ke mamalia, namun karena ada NNP di kalimat yang sama membuat subjek asli yang dikenali adalah kelenjar susu
Itu	Mereka terus memantau perang Pasifik melalui radio-radio gelap. Sebab pada saat itu Jepang melarang bangsa Indonesia memiliki pesawat komunikasi.	Pronomina “itu” tidak mendeskripsikan subjek yang secara eksplisit ada pada kalimat sebelumnya

Hasil evaluasi menunjukkan bahwa aturan yang diusulkan dapat mengenali pronomina dengan akurasi lebih dari 80%. Rata-rata akurasi yang didapat adalah 81%. Metode pengenalan pronomina menggunakan aturan ini memang kurang efektif jika dibandingkan dengan metode *machine learning*. Namun, metode *machine learning* membutuhkan data latih yang cukup banyak dan saat ini data latih tersebut belum tersedia untuk versi Bahasa Indonesia.

Oleh sebab itu, penelitian awal ini dirancang agar dapat digunakan untuk membentuk ratusan

bahkan ribuan dataset berupa kombinasi kalimat pronomina dan kalimat tanpa pronomina. Kemudian, dataset tersebut dapat digunakan sebagai data latih untuk proses pengenalan pronomina menggunakan metode lain dalam *machine learning*.

4. KESIMPULAN

Proses pengenalan pronomina diperlukan untuk memperjelas konteks suatu kalimat. Proses ini dapat diselesaikan dengan menggunakan metode berbasis aturan. Aturan tersebut dibentuk berdasarkan jenis pronomina yang digunakan dan struktur kalimat yang menyertainya. Berdasarkan proses percobaan dan evaluasi yang telah dilakukan, diketahui bahwa metode yang diusulkan telah berhasil mengenali pronomina. Rata-rata akurasi yang didapat adalah 81%. Meskipun mendapatkan akurasi yang tinggi, beberapa studi kasus masih belum dapat terselesaikan. Salah satu penyebab kegagalan pengenalan pronomina adalah adanya kalimat yang terlalu kompleks dan tidak tersedianya pola yang sesuai. Oleh karena itu, penelitian selanjutnya akan mencoba mengenali pronomina menggunakan metode lain agar pembelajaran pola bisa lebih efektif.

REFERENCES

- ALORAINI, A. & POESIO, M. 2020. Anaphoric Zero Pronoun Identification : A Multilingual Approach’, in *Proceedings of the 3rd Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2020)*, pp. 22–32.
- ANITIA, D., MUNARKO, Y. & AZHAR, Y. 2020. Parsing Twitter Menggunakan Metode Left-Corner Parsing Dengan Memanfaatkan Pos Tagger. *Jurnal Repositor*, 2(7), p. 897. doi: 10.22219/repositor.v2i7.203.
- AWASTHI, I. *dkk.* 2021. Natural Language Processing (NLP) based Text Summarization - A Survey’, in *Sixth International Conference on Inventive Computation Technologies*, pp. 1310–1317.
- BASUKI, S. & KUSUMA, S. F. 2018. Automatic question generation for 5W-1H open domain of Indonesian questions by using syntactical template-based features from academic textbooks. *Journal of Theoretical and Applied Information Technology*, pp. 3908–3923.
- CHANG, T. *dkk.* 2017. Zero Pronoun Identification in Chinese Language with Deep Neural Networks’, in *2nd International Conference on Control, Automation, and Artificial Intelligence (CAAI 2017)*, pp. 518–522.
- KASTRATI, Z. *dkk.* 2021. Sentiment Analysis of Students’ Feedback with NLP and Deep Learning: A Systematic Mapping Study’, *mdpi*.
- KHOIRUNISA, R. *dkk.* 2020. Penggunaan Natural

- Language Processing Pada Chatbot Untuk Media Informasi Pertanian. *Indonesian Journal of Applied Informatics*, 4(2), p. 55. doi: 10.20961/ijai.v4i2.38688.
- KUSUMA, S. F. *dkk.* 2018. Indonesian Question Generation Based on Bloom's Taxonomy Using Text Analysis', in *Proceeding - 2018 International Seminar on Intelligent Technology and Its Application, ISITIA 2018*. IEEE, pp. 269–274. doi: 10.1109/ISITIA.2018.8711015.
- KUSUMA, S. F. & ALHAMRI, R. Z. 2018. Generating Indonesian Question Automatically Based on Bloom's Taxonomy Using Template Based Method', *KINETIK: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 3(2), pp. 145–152. doi: 10.22219/kinetik.v3i2.650.
- KUSUMA, S. F., SUKYA, F. & HERIADI, A. 2021. Pendekatan Baru untuk Merepresentasi Informasi di Bidang Pendidikan Menggunakan Kombinasi Ontologi. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 7(2), p. 160. doi: 10.26418/jp.v7i2.46978.
- MOELIONO, A. M. *ET AL.* 2017. *Adan pengembang d pe b kementerian pendidikan dan kebud.*
- PISTOL, I., TRANDABÄT, D. & RÄSCHIP, M. 2018. Medi-Test: Generating Tests from Medical Reference Texts', *mdpi*. doi: 10.3390/data3040070.
- PRATAMA, M. R., KUSUMADEWI, S. & HIDAYAT, T. 2017. Penerapan algoritma. *Jurnal Teknologi Elektro, Universitas Mercu Buana*, 8(1), pp. 1–8.
- SARIPUDIN, P. & PURNAMASARI, K. K. 2017. Pendeteksian Keterkaitan Antar Kalimat Dengan Metode Template - Based Dalam Pembangkit Pertanyaan Otomatis', pp. 1–7.
- SUGONO. 2008. *Kamus Bahasa Indonesia*. Jakarta: Pusat Bahasa Departemen Pendidikan Nasional.
- WARDANA, H. K., SWANITA, I. & YOHANES, B. W. 2019. Sistem Pemeriksa Pola Kalimat Bahasa Indonesia berbasis Algoritme Left-Corner Parsing dengan Stemming. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 8(3), p. 211. doi: 10.22146/jnteti.v8i3.515.

Halaman ini sengaja dikosongkan