

REVIEW JURNAL PRESENTATION

Detecting Indonesian ambiguous sentences
using Boyer Moore

ANITA DWI MAHARANI

2321210007



Detecting Indonesian ambiguous sentences using Boyer Moore

Jurnal : TELKOMNIKA, Telecommunication Computing Electronics and Control

Volume dan Halaman : 18, No. 5, October 2020, pp. 2480~2487

Author

- Risky Aswi Ramadhani,
- I Ketut Gede Darma Putra,
- Made Sudarma,
- I. A. D. Giriantari

LATAR BELAKANG

Penelitian yang dibahas penulis berkaitan masalah **kalimat ambigu**, yang dapat diklasifikasikan menjadi ambiguitas fonetik, leksikal, dan tata bahasa.

Secara khusus, fokusnya adalah pada ambiguitas tata bahasa yang dihasilkan dari penggunaan **tata bahasa yang salah**. Kalimat yang memiliki banyak makna dikenal sebagai kalimat ambigu.

Fokus penelitian ini adalah pada ambiguitas tata bahasa. Penggunaan tata bahasa yang salah menyebabkan **ambiguitas tata bahasa**. Namun, ambiguitas ini cenderung menghilang ketika kalimat digunakan. Penelitian tentang kalimat yang ambigu secara tata bahasa, terutama **deteksi** mereka dalam bahasa Indonesia, **masih belum dieksplorasi**.



LATAR BELAKANG

“bulan (bulan/bulan)”

Dengan menggunakan sistem untuk mendeteksi kalimat yang ambigu secara tata bahasa, menjadi mungkin untuk memahami dan menerjemahkan makna yang dimaksudkan dari kalimat tersebut. Tujuannya adalah untuk meningkatkan pemahaman komputer tentang kalimat ambigu dalam bahasa Indonesia. Studi yang ada telah **mengidentifikasi kalimat yang ambigu** tetapi belum sepenuhnya memahami artinya.

Untuk **meningkatkan akurasi sistem terjemahan** dan memfasilitasi pemahaman komputer tentang konten tekstual. Oleh karena itu, hal baru yang dicari dalam penelitian ini adalah pengembangan sistem deteksi kalimat ambigu tata bahasa dalam bahasa Indonesia menggunakan algoritma BoyerMoore.



TUJUAN

Dengan memanfaatkan algoritma Boyer-Moore dan metode persamaan kosinus, penelitian ini berupaya **meningkatkan keakuratan mendeteksi kalimat ambigu** dan menentukan **artinya**, sehingga **memfasilitasi pemahaman teks oleh komputer dan meningkatkan sistem terjemahan**

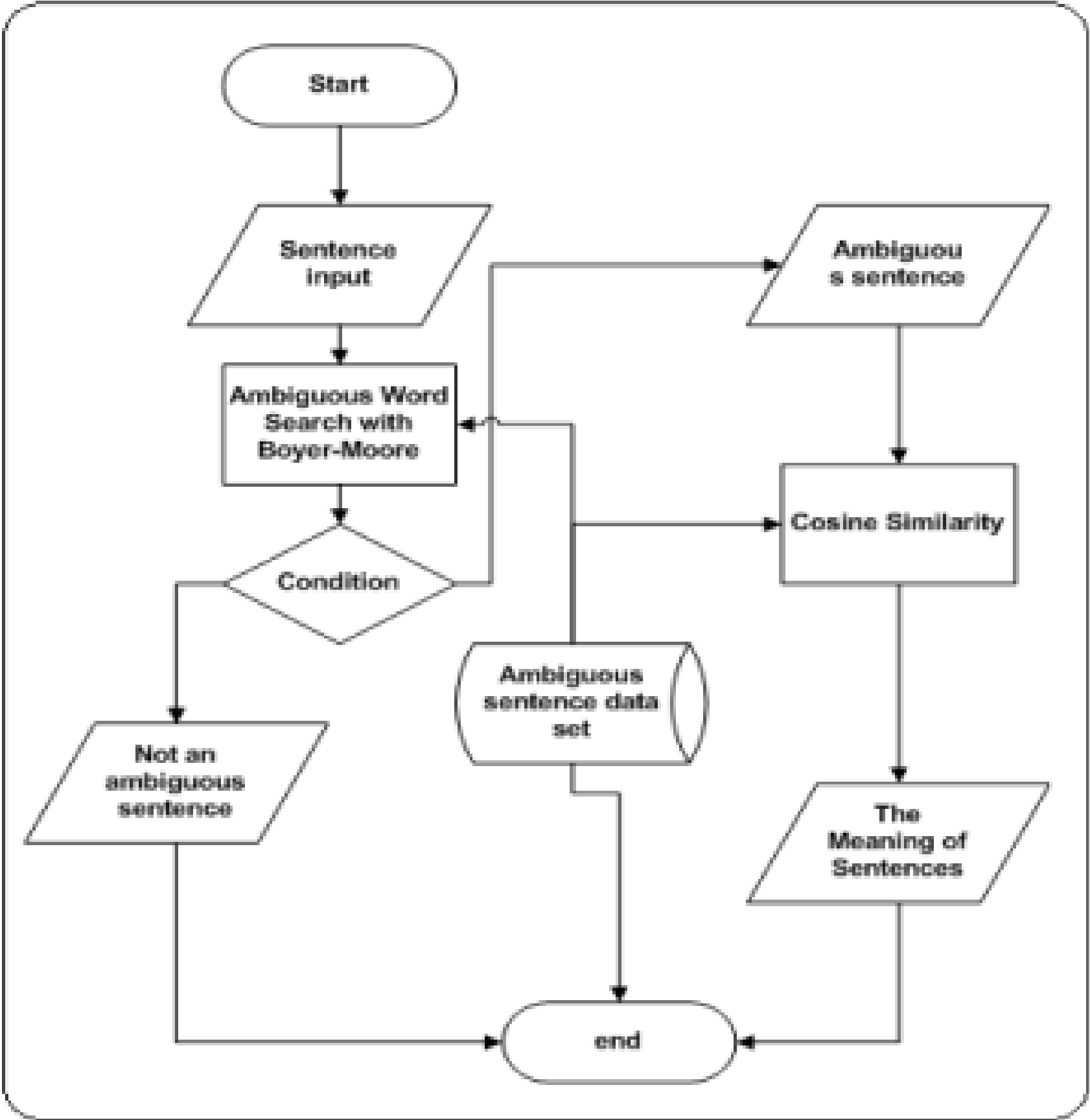


METODE PENELITIAN

Dengan memanfaatkan **algoritma Boyer-Moore** dan **metode persamaan kosinus**, penelitian ini berupaya meningkatkan keakuratan mendeteksi kalimat ambigu dan menentukan artinya, sehingga memfasilitasi pemahaman teks oleh komputer dan meningkatkan sistem terjemahan

penelitian ini melakukan **200 percobaan** untuk menguji kinerja sistem dalam mendeteksi kalimat ambigu tata bahasa menggunakan algoritma Boyer-Moore dan metode persamaan kosinus. Dengan menggunakan **data set 50** kata ambigu, kalimat ambigu, dan makna kalimat

Kemudian kinerja sistem dievaluasi juga menggunakan **confusion matrix**, metode umum untuk menilai presisi, akurasi, dan recall.



DATA SET

Table 1. Ambiguous words and sentences

Ambiguous words	Sentence	Meaning
<i>Budi (Mind)</i>	<i>Aku mengenang budi baikmu (I remember your kindness)</i>	<i>Kebaikan (Kindness)</i>
<i>Salam (Regards)</i>	<i>Gus kamu kemarin mendapatkan salam dari anggi (Gus, Anggi sent you regards yesterday)</i>	<i>Sapaan (Greetings)</i>
<i>Tahu (Tofu)</i>	<i>Agus kesini tadi memberi tahu (Agus came here to give us tofu)</i>	<i>Makanan (Food)</i>
<i>Bunga (Interest)</i>	<i>Bunga deposito di bank jatim lumayan tinggi (The deposit interest rate in Bank Jatim is quite high)</i>	<i>Keuntungan (Profit)</i>
<i>Bangku (Bench)</i>	<i>Dia tidak pernah makan bangku sekolah (He never went to school)</i>	<i>Pendidikan (Education)</i>
<i>Kemas (Organized)</i>	<i>Acara ini dikemas dengan sangat baik (This event is very well organized)</i>	<i>Melakukan pekerjaan (Doing work)</i>
<i>Bulan (Month)</i>	<i>Awal Bulan Kamu gajian (You are paid at the beginning of the month)</i>	<i>Waktu (Time)</i>

HASIL PENELITIAN

“Setiap awal bulan kami gajian (We are paid at the beginning of each month)”

- Bulan (month) = a period of time
- Bulan (moon) = sky object

Menemukan Kata Ambigu menggunakan
Algoritma Boyer Moore

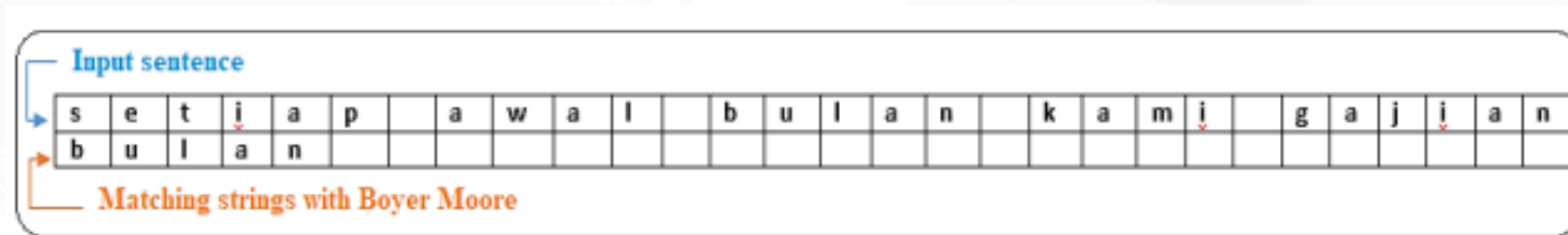


Figure 3. Ambiguous word search step 1

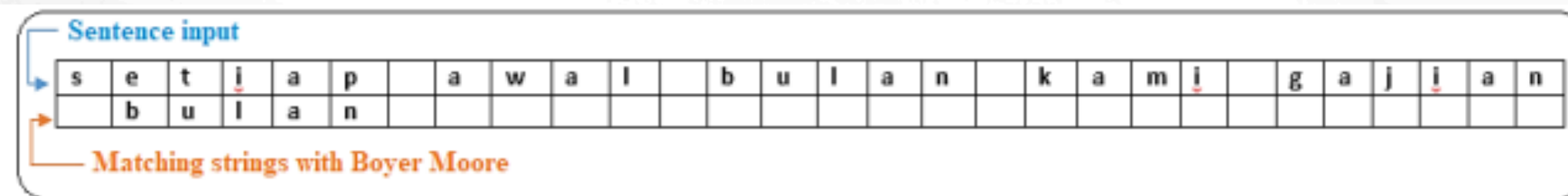


Figure 4. Ambiguous word search step 2

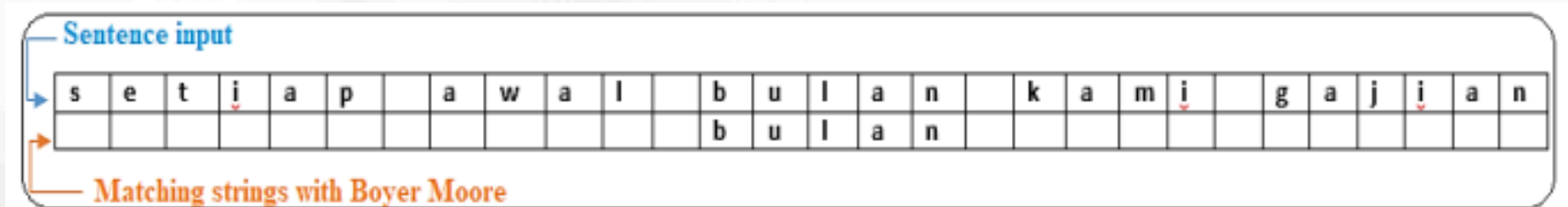


Figure 5. Ambiguous word search step 13

HASIL PENELITIAN

Perbandingan Kalimat Data Set

- Awal bulan kami gajian (We are paid at the beginning of the month).
- Bumi dan bulan merupakan benda langit (The Earth and The Moon are sky objects).

Menemukan Kesamaan Kata

Cosine Similarity

- S1 = awal bulan kami gajian (We are apid at the beginning of the month).
- S2 = awal bulan kamu gajian (You are paid at the beginning of the month)

Table 2. Ambiguous words and sentences

Word	Count		A. B	A ²	B ²
	A	B			
<i>Awal</i> (beginning)	1	1	1	1	1
<i>Bulan</i> (month)	1	1	1	1	1
<i>Kami</i> (we)	1	0	0	1	0
<i>Kamu</i> (you)	0	1	0	0	1
<i>Gajian</i> (paid)	1	0	0	1	0
			2	4	3

$$\begin{aligned} \text{Similarity} = \cos(\theta) &= \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{2}{\sqrt{4 \times 3}} \\ &= 0.166 \end{aligned}$$

HASIL PENELITIAN

Id	Input Sentences	Data Set	Value of Similarity
1	<i>Setiap awal bulan kami gajian (We are paid at the beginning of each month)</i>	<i>Setiap awal bulan kamu gajian (You are paid at the beginning of each month)</i>	0.16
2	<i>Setiap awal bulan kami gajian (We are paid at the beginning of each month)</i>	<i>Bumi dan bulan merupakan benda langit (Earth and moon are sky objects)</i>	0.05

Ambiguous sentence detection results

ambiguous words	Cut string	significance	cosine similarity value
Bulan	Awal bulan kamu	waktu	0,16
Bulan	Bumi bulan merupakan	benda	0,05

HASIL PENELITIAN

Untuk analisis hasil tes dari 200 uji percobaan dari 50 data set, menggunakan Confusion Matrix value, kemudian menghitung accuracy, precision, recall, dan F-measure

Table 4. Confusion matrix value

TP=40	FP=3
FN=10	TN=147

$$Accuracy = \frac{40+187}{40+147+3+10} = 0.935$$

$$Precision = \frac{40}{40+3} = 0.9302$$

$$Recall = \frac{40}{40+3} = 0.8$$

Sistem mencapai tingkat **akurasi** 0,935, **presisi** 0,9320, dan **recall** 0,8, menunjukkan kemampuannya untuk mengenali kata-kata ambigu secara efektif. Meskipun dengan keterbatasan data set yang digunakan.

F-Measure, metrik yang menggabungkan metode umum tersebut dihitung menghasilkan 0,8061, menunjukkan kinerja keseluruhan sistem dalam mengidentifikasi kalimat ambigu.

Kecepatan rata-rata untuk pencarian kata adalah 0,003275 detik, menunjukkan efisiensi sistem dalam memproses kalimat ambigu.

Studi ini berhasil menggunakan algoritma Boyer-Moore dan metode persamaan kosinus untuk mendeteksi dan menafsirkan kata-kata ambigu dalam kalimat Indonesia, menyoroti efektivitas sistem dalam memahami ambiguitas tata Bahasa.

KELEBIHAN

Kebaruan penelitian ini terletak pada fokusnya pada pengembangan sistem deteksi untuk kalimat ambigu tata bahasa Indonesia menggunakan algoritma Boyer-Moore, area yang kurang dieksplorasi dalam penelitian.

Tidak seperti penelitian lainnya, seperti yang diakui penulis, yang hanya mengidentifikasi kalimat ambigu tanpa menafsirkan artinya, penelitian ini melangkah lebih jauh dengan memanfaatkan metode kesamaan kosinus untuk memahami konteks dan makna kata-kata ambigu dalam kalimat.

Dengan mengatasi tantangan ambiguitas tata bahasa dalam kalimat bahasa Indonesia, penelitian ini berkontribusi untuk **meningkatkan akurasi dan efisiensi sistem pemrosesan bahasa, seperti penerjemah, dengan memungkinkan komputer untuk lebih memahami dan menafsirkan teks ambigu**

KEKURANGAN

penulis mengakui kurangnya kumpulan data yang tersedia yang mencakup kalimat ambigu tata bahasa Indonesia, yang menimbulkan tantangan dalam melatih dan menguji sistem deteksi secara efektif .

Keterbatasan ini dapat memengaruhi kemampuan sistem untuk mengenali berbagai kata dan kalimat ambigu yang lebih luas secara akurat.

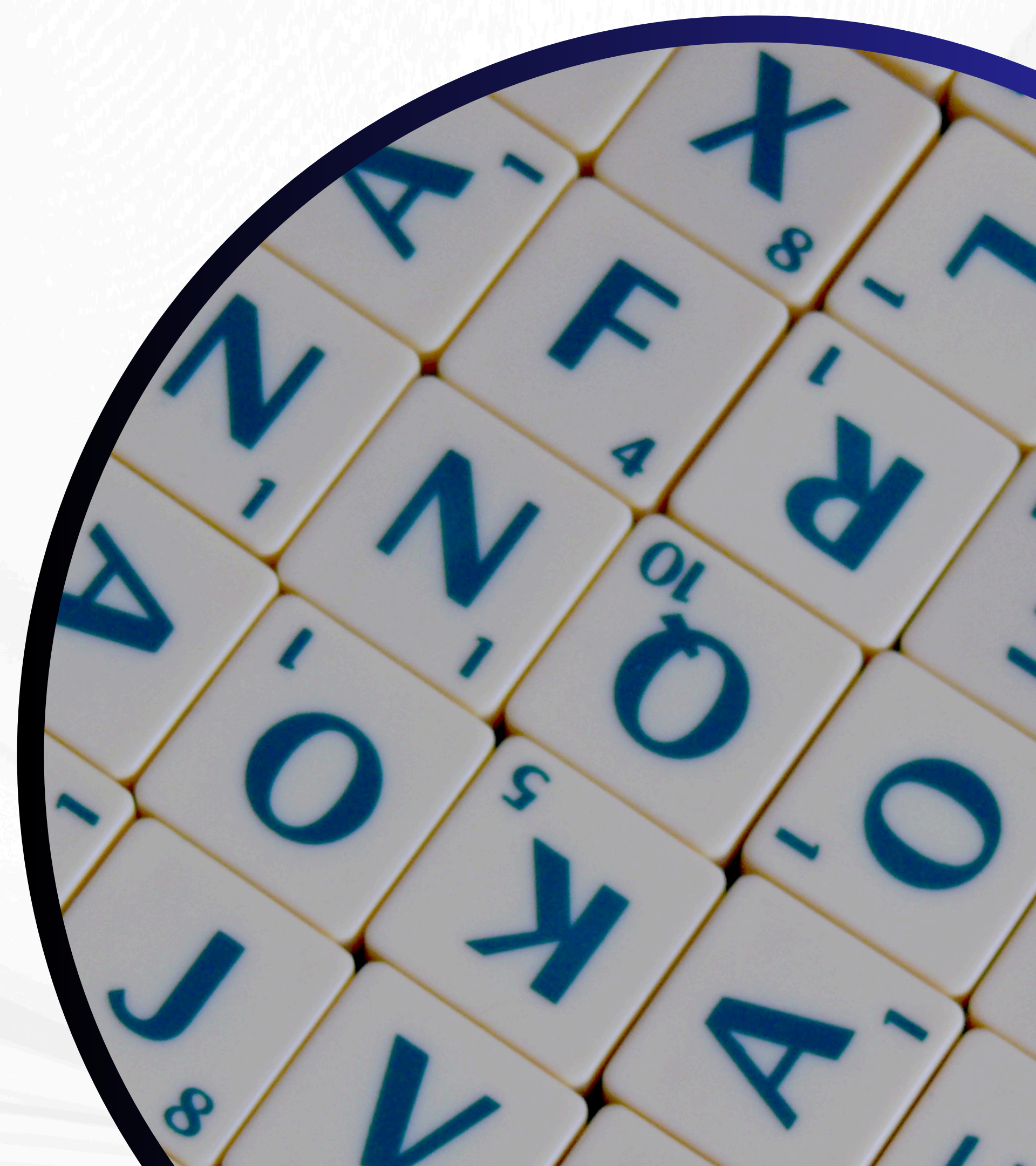
Keterbatasan ini menunjukkan potensi penyempurnaan dan peningkatan lebih lanjut dalam mendeteksi dan menafsirkan konstruksi bahasa yang ambigu.

KESIMPULAN

Penelitian menyimpulkan bahwa kalimat ambigu tata bahasa Indonesia, yang memiliki dua makna, dapat dikenali secara efektif menggunakan algoritma Boyer-Moore dan persamaan kosinus.

Algoritma Boyer-Moore berperan penting dalam menemukan string ambigu dalam kalimat, sedangkan algoritma persamaan kosinus menghitung tingkat kesamaan antara objek linguistik yang berbeda, membantu dalam menentukan makna kalimat.

Studi ini menyoroti efektivitas algoritma kesamaan Boyer Moore dan kosinus dalam mendeteksi kata-kata ambigu, sebagaimana dibuktikan oleh tingkat recall sistem yang tinggi sebesar 80% dan kecepatan rata-rata 0,003275 detik untuk mendeteksi kalimat ambigu.





THANKS FOR WATCHING

www.reallygreatsite.com

