

Sistem Deteksi Kesalahan Tanda Baca dan Huruf Kapital Pada Karya Tulis Ilmiah Berbahasa Indonesia Menggunakan Algoritma *Boyer-Moore*

Mazidhatul Ilmiyah¹, Anita Qoiriah²

^{1,2}Jurusan Teknik Informatika/Teknik Informartika, Universitas Negeri Surabaya

¹Mazidhatulilmiyah16051204011@mhs.unesa.ac.id

²anitaqoiriah@unesa.ac.id

Abstrak—Karya tulis ilmiah adalah karangan yang dipublikasikan atau dipaparkan dari hasil pemantauan, pengamatan, serta peninjauan dalam bidang dan menggunakan metode tertentu serta sistematika penulisan yang baik. Serta apa saja yang dipaparkan bisa dipertanggungjawabkan keilmiahannya. Karya tulis ilmiah yang baik adalah karya tulis yang menggunakan bahasa Indonesia dengan baik dan sesuai aturan, entah itu dari segi tata bahasa atau dalam penulisannya. Karena hal tersebut harus sesuai dengan bidangnya, yaitu ragam keilmuan. Dalam penulisan karya tulis ilmiah perlu memperhatikan dalam penggunaan tanda baca, pilihan kata, dan kesesuaian dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). PUEBI berisi tentang aturan penggunaan Bahasa Indonesia yang baik dan benar. Kesalahan penulisan Bahasa Indonesia yang banyak ditemukan adalah kesalahan dalam penggunaan tanda baca dan penulisan huruf kapital yang benar. Maka dari itu perlu untuk melakukan penelitian guna menentukan solusi permasalahan dalam mendeteksi kesalahan penulisan huruf kapital dan penggunaan tanda baca dalam suatu dokumen. Pada penelitian ini, penulis menggunakan algoritma *Boyer-Moore* yang terbukti efisien sebagai algoritma untuk aplikasi pencarian kata atau *string*. Algoritma ini banyak diimplementasikan sebagai fitur “*search*”. Penelitian ini menghasilkan bukti bahwa algoritma *Boyer-Moore* bisa mendeteksi kesalahan dengan nilai rata-rata presisi sebesar 0,969, untuk nilai rata-rata *recall* sebesar 0,976, dan untuk nilai rata-rata akurasi sebesar 0,917.

Kata Kunci— karya tulis ilmiah, tanda baca, huruf kapital, *Boyer Moore*, Pencarian *string*.

I. PENDAHULUAN

Kegiatan menulis adalah keterampilan dasar seseorang dalam berbahasa, menulis termasuk salah satu sarana untuk menyalurkan ide dan gagasan oleh penulis. Maka dari itu penulis harus memilih kata, menyusun tujuan, lalu menuangkannya dalam bahasa yang mudah dimengerti oleh pembacanya. Kegiatan menulis salah satu fungsinya adalah sebagai alat komunikasi secara tidak langsung. Beberapa fungsi lain dari menulis yaitu mempertajam daya persepsi, membantu untuk berpikir lebih kritis, membantu memecahkan masalah, serta dapat membantu dalam menjelaskan pikiran-pikiran.[1]

Salah satu hasil dari kegiatan menulis adalah berkarya dalam menciptakan suatu karya tulis ilmiah. Karya tulis ilmiah adalah suatu tulisan atau karangan yang dipublikasikan atau dipaparkan berdasarkan pemantauan, pengamatan, penelitian, serta peninjauan dalam bidang tertentu yang dapat dipertanggungjawabkan keilmiahannya. Adapun syarat keilmiahannya suatu karya ilmiah yaitu menyajikan fakta secara objektif, cermat, dan jujur. Penyusunannya dilakukan secara sistematis, serta menyajikan penalaran sebab-akibat, mengandung pandangan peneliti, dan tanpa manipulasi.[2]

Dalam menulis karya tulis ilmiah yang harus diperhatikan adalah aturan penggunaan tanda baca, pilihan kata dan kesesuaian dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). PUEBI berisi tentang aturan penggunaan Bahasa Indonesia yang baik dan benar. Pada tanggal 26 November 2015 PUEBI secara resmi diterapkan oleh menteri pendidikan saat itu, yaitu Dr. Anis Baswedan. PUEBI berisi petunjuk atau pedoman dalam pemakaian ejaan bahasa Indonesia, antara lain berisi aturan dalam penulisan kata, pemakaian huruf, serta aturan dalam pemakaian tanda baca.

Adanya aturan PUEBI diharapkan para peneliti atau pencipta suatu karya tulis ilmiah dapat meminimalisir kesalahan dalam hal penulisan. Kesalahan penulisan Bahasa Indonesia yang banyak terjadi adalah kesalahan menulis huruf kapital dan juga kesalahan dalam menggunakan tanda baca. Kesalahan seperti ini membuat makna ambigu bagi pembaca, juga dapat mengubah arti dan makna dari suatu kalimat. Maka dari itu untuk menentukan solusi dalam permasalahan banyak peneliti yang terdorong untuk melakukan penelitian. Salah satu penelitian yang berkaitan dengan penggunaan tanda baca adalah “Kakas Bantu Pendeteksi Kesalahan Tanda Baca pada Karya Tulis Ilmiah” oleh Mohammad Ahmaluddin Zinni, Ratih Nur Esti Anngraini, dan Siti Rochimah pada tahun 2016 dengan menggunakan algoritma *Boyer-Moore*. Dari 5 data dokumen uji coba mendapatkan hasil nilai rata-rata presisi sebesar 0,6806, nilai rata-rata *recall* sebesar 0,969 dan untuk nilai rata-rata akurasi sebesar 0,9636. Penelitian tersebut membuktikan bahwa algoritma *Boyer-Moore* dapat digunakan untuk pencarian tanda baca, hanya saja pada penelitian ini belum dilengkapi dengan deteksi kesalahan huruf kapital pada karya tulis ilmiah.[3]

Ada banyak contoh algoritma pencarian kata diantaranya algoritma *Knuth-Morris-Pratt*, *Rabin-Karp*, *Boyer Moore*, dll. Pada penelitian perbandingan algoritma pencarian kata dengan judul “Perbandingan Algoritma *Knuth Morris Pratt* dan *Boyer*

Moore dalam Pencocokan *String* pada Aplikasi Kamus Bahasa Nias” oleh Rila Mandala dan Fince Tinus Waruwu pada tahun 2016, mendapatkan hasil bahwa kedua algoritma tersebut dapat diterapkan di dalam aplikasi kamus Nias agar mempermudah proses pencarian dan pencocokan karakter kamus Nias. Dan jika dibandingkan antara kedua algoritma tersebut bahwa algoritma *Boyer Moore* lebih efisien dalam waktu akses pencarian kata[4]

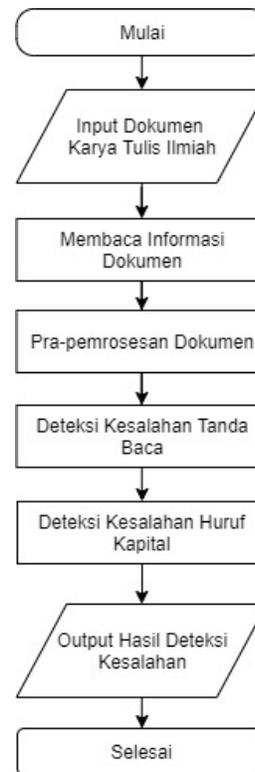
Penelitian penggunaan algoritma *Boyer-Moore* dilakukan oleh Agus Pratama dan Rohmat Indra Borman pada tahun 2016 dengan judul “Penerapan *String Matching* dengan algoritma *Boyer-Moore* pada Aplikasi *Font Italic* untuk Deteksi Kata Asing”. Hasil dari penelitian ini adalah pembuktian bahwa algoritma *Boyer-Moore* dapat digunakan untuk pencocokan *string* pada pencarian kata berbahasa Inggris. Hasil uji proses deteksi kata berbahasa Inggris sudah sesuai dengan kata yang berada di database, namun penelitian ini memiliki kekurangan yakni apabila kata yang ada pada teks semakin banyak, maka waktu yang dibutuhkan dalam pencarian juga semakin bertambah.[5]

Pada penelitian algoritma *Boyer-Moore* lain yang dilakukan oleh Ansari Saleh Ahmar, Robbi Rahim, Dicky Nofriansyah, dan Ayu Putri Ardyanti pada tahun 2017 dengan judul “*Visual Approach of Searching Process using Boyer-Moore Algorithm*” yang melakukan simulasi pecocokan kata menggunakan algoritma *Boyer-Moore* menunjukkan hasil bahwa algoritma tersebut merupakan algoritma pencarian kata yang dapat diimplementasikan ke dalam banyak proses pencarian.[6]

Berdasarkan latar belakang permasalahan diatas, peneliti ingin membuat skripsi berjudul “Sistem Deteksi Kesalahan Tanda Baca dan Huruf Kapital pada Karya Tulis Ilmiah Menggunakan Algoritma *Boyer-Moore*”. Penulis berharap sistem ini dapat membantu para penulis karya tulis ilmiah berbahasa Indonesia dalam mendeteksi kesalahan penulisan tanda baca dan huruf kapital.

II. METODOLOGI PENELITIAN

Tujuan dari penelitian ini untuk menerapkan metode *text-mining* pencarian kata yaitu *Boyer-Moore* sebagai dasar untuk mendeteksi kesalahan penggunaan tanda baca dan huruf kapital pada karya tulis ilmiah berbahasa Indonesia. Pada penelitian ini terdapat beberapa tahapan untuk melakukan penelitian. Gambaran proses dari sistem ditunjukkan pada Gbr. 1



Gbr. 1 Alur Sistem

Data yang dimasukkan ke dalam program berupa karya tulis ilmiah berbahasa Indonesia dengan format .docx, Setelah data dimasukkan, sistem akan membaca file dokumen tersebut, kemudian dilakukan tahap pra-pemrosesan. Setelah tahap pra-pemrosesan selesai, maka sistem akan mulai mendeteksi kesalahan penggunaan tanda baca dan huruf kapital menggunakan algoritma *Boyer-Moore*.

A. Karya Tulis Ilmiah

Pada hakikatnya, karya tulis ilmiah madalah suatu tulisan atau karangan yang dipublikasikan atau dipaparkan dari hasil pantauan, penelitian, serta pengamatan dalam bidang dan metode tertentu serta sistematika dalam penulisan yang baik, dan juga keilmiahan dari isinya dapat dipertanggungjawabkan.

Karya tulis ilmiah adalah hasil penelitian yang dilakukan oleh individu maupun tim yang memenuhi kaidah keilmuan. Akan tetapi, karya tulis ilmiah ada juga yang diperoleh dari hasil kajian pustaka seperti jurnal, makalah, dll.

Beberapa hal harus diperhatikan dalam menulis karya tulis ilmiah, yakni dalam penggunaan tanda baca, pemilihan kata, dan kesesuaian dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI) adalah hal yang perlu diperhatikan.[7]

PUEBI berisi petunjuk atau pedoman dalam pemakaian ejaan bahasa Indonesia, antara lain berisi aturan dalam

penulisan kata, pemakaian huruf, serta aturan pemakaian tanda baca.

B. Huruf Kapital

Aturan penggunaan huruf di dalam Pedoman Umum Bahasa Indonesia (PUEBI) ada 2 jenis, yakni aturan dalam penulisan huruf besar atau huruf kapital dan penulisan huruf miring. Terdapat 15 aturan untuk pemakaian huruf kapital. Aturan huruf kapital yang digunakan pada penelitian ini adalah:

1. Huruf kapital pada awal kalimat.

Contoh:

Pada zaman dahulu tinggal seekor kancil di hutan.

2. Huruf kapital pada nama orang.

Contoh:

Nur Lujeng Kinanti

3. Huruf kapital pada nama negara

Contoh:

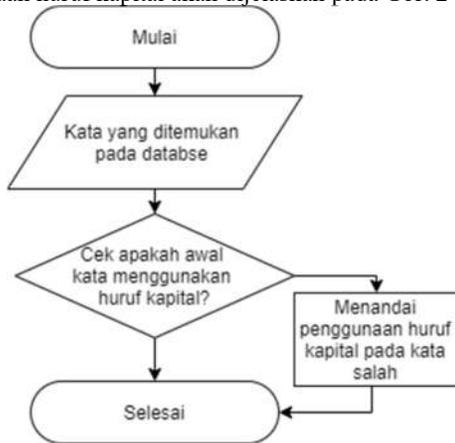
Saya lahir di Indonesia.

4. Huruf kapital pada keterangan waktu.

Contoh:

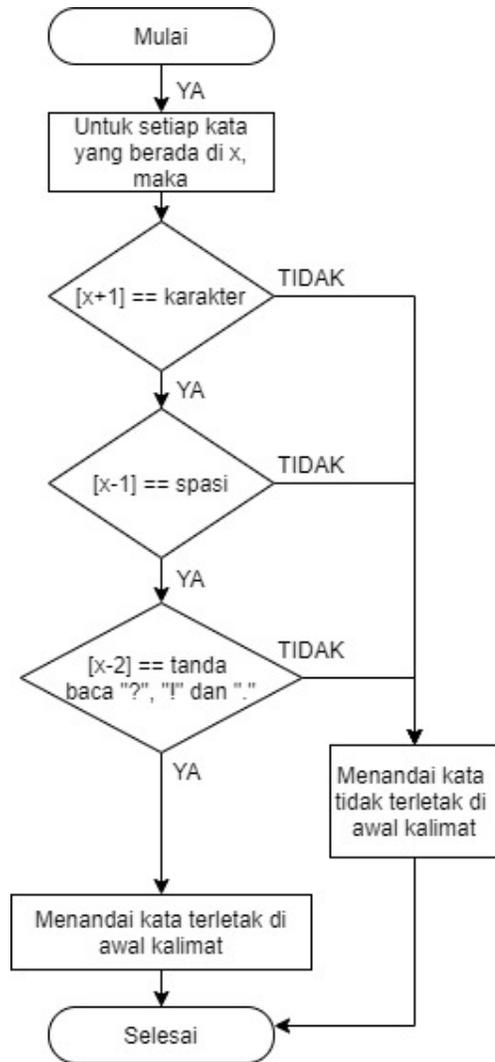
Aku berencana pergi di Hari Minggu.

Untuk proses pencarian kesalahan penggunaan huruf kapital, akan dilakukan pencarian kata menggunakan algoritma pencarian kata *Boyer-Moore*. Alur untuk pemeriksaan huruf kapital akan dijelaskan pada Gbr. 2



Gbr. 2 Alur Pemeriksaan Huruf Kapital

Pada penelitian ini, menggunakan 3 korpus data yang digunakan sebagai acuan pemeriksaan, diantaranya data nama orang, nama negara, dan data keterangan waktu. Dari ke-3 korpus tersebut, akan dilakukan pencocokan kata dengan kata yang ada pada dokumen uji, apabila ditemukan maka akan diperiksa kebenaran penggunaan huruf kapitalnya. Sedangkan untuk pemeriksaan huruf kapital pada awal kalimat, menggunakan alur khusus yang akan ditunjukkan pada Gbr 3.



Gbr. 3 Alur Pemeriksaan Huruf Kapital Awal Kalimat

C. Tanda Baca

Tanda-tanda yang digunakan pada ejaan terutama yang berbentuk tulisan disebut tanda baca. Fungsi dari tanda baca yakni sebagai aturan dan suara pembicara dalam memberi arti terhadap suatu kalimat penjelasan.

Aturan penggunaan tanda baca akan dijelaskan pada tabel 1.

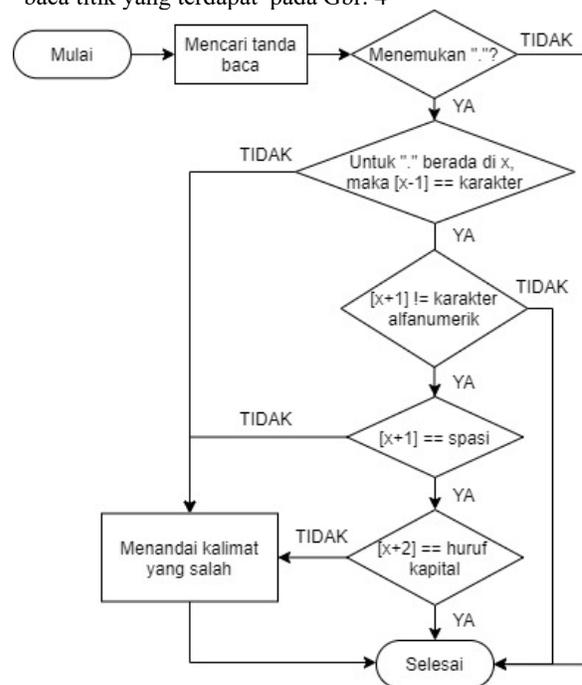
TABEL I
 ATURAN TANDA BACA

No	Jenis Tanda Baca	Aturan Penggunaan
1.	Titik (.)	Sebelum tanda baca titik “.” harus didahului sebuah karakter. Setelah tanda baca titik “.” tidak diikuti dengan penggunaan spasi jika karakter selanjutnya alfanumerik. Setelah tanda baca titik “.” diikuti dengan

		penggunaan spasi jika karakter selanjutnya adalah karakter alfabet. Setelah tanda baca titik “.” dan spasi harus diikuti penggunaan huruf kapital.
2.	Koma (,)	Sebelum tanda baca koma “,” harus didahului sebuah karakter. Setelah tanda baca koma “,” tidak diikuti dengan penggunaan spasi jika karakter selanjutnya bukan alfanumerik. Setelah tanda baca koma “,” diikuti dengan penggunaan spasi jika karakter selanjutnya adalah alfabetik. Setelah tanda baca koma “,” dan spasi harus diikuti penggunaan huruf kecil (bukan kapital)
3.	Titik Koma (;)	Sebelum tanda baca titik koma “;” harus didahului sebuah karakter. Setelah tanda baca titik koma “;” diikuti dengan penggunaan spasi. Setelah tanda baca titik koma “;” dan spasi harus diikuti penggunaan huruf kecil (bukan kapital)
4.	Titik Dua (:)	Sebelum tanda baca titik dua “:” harus didahului sebuah karakter. Setelah tanda baca titik dua “:” diikuti dengan penggunaan spasi. Setelah tanda baca titik dua “:” dan spasi harus diikuti penggunaan huruf kecil (bukan kapital).
5.	Tanda Hubung (-)	Sebelum tanda baca hubung “-” harus didahului sebuah karakter. Setelah tanda baca hubung “-” tidak dapat diikuti dengan penggunaan spasi. Setelah tanda baca hubung “-” harus diikuti penggunaan huruf kecil (bukan kapital)
6.	Tanda Tanya (?)	Sebelum tanda tanya “?” harus didahului sebuah karakter. Setelah tanda tanya “?” diikuti dengan penggunaan spasi. Setelah tanda tanya “?” dan spasi harus diikuti dengan penggunaan huruf kapital.
7.	Tanda Seru (!)	Sebelum tanda seru “!” harus didahului sebuah karakter. Setelah tanda seru “!” diikuti dengan penggunaan spasi. Setelah tanda seru “!” dan spasi harus diikuti penggunaan huruf kapital.
8.	Tanda Kurung Buka (())	Sebelum tanda baca kurung buka “(“ harus didahului spasi. Setelah tanda baca kurung buka “(“ tidak dapat diikuti dengan penggunaan spasi. Setelah tanda baca kurung buka “(“ harus dilengkapi dengan adanya tanda kurung tutup “)” dalam satu kalimat.
9.	Tanda Kurung Tutup ())	Sebelum tanda baca kurung tutup “)” harus didahului sebuah karakter. Setelah tanda baca kurung tutup “)” harus diikuti penggunaan spasi jika tidak ada titik pada

		karakter selanjutnya. Sebelum tanda baca kurung tutup “)” harus didahului dengan adanya tanda kurung buka “(“ dalam satu kalimat.
10.	Tanda Kurung Siku Buka ([])	Sebelum tanda baca kurung siku buka “[” harus didahului spasi. Setelah tanda baca kurung siku buka “[” tidak dapat diikuti dengan penggunaan spasi. Setelah tanda baca kurung siku buka “[“ harus dilengkapi dengan adanya tanda kurung siku tutup “]” dalam satu kalimat.
11.	Tanda Kurung Siku Tutup (])	Sebelum tanda baca kurung siku tutup “]” harus didahului sebuah karakter. Setelah tanda baca kurung siku tutup “]” harus diikuti penggunaan spasi jika tidak ada titik pada karakter selanjutnya. Sebelum tanda baca kurung siku tutup “]” harus didahului dengan adanya tanda kurung siku buka “[“ dalam satu kalimat.
12.	Garis Miring (/)	Sebelum tanda garis miring “/” harus didahului sebuah karakter. Setelah tanda garis miring “/” tidak dapat diikuti dengan penggunaan spasi. Setelah tanda garis miring “/” harus diikuti penggunaan huruf kecil (bukan kapital).

Berikut ini contoh salah satu alur pemeriksaan tanda baca titik yang terdapat pada Gbr. 4



Gbr. 4 Alur Pemeriksaan Tanda Baca Titik

D. Algoritma Boyer-Moore

Pada pencarian kata, algoritma *Boyer-Moore* merupakan salah satu algoritma yang terbukti efisien.[8]

Algoritma ini banyak diimplementasikan sebagai fitur “*search*”. Algoritma ini cara kerjanya hampir sama dengan algoritma *Knuth-Morris-Pratt*, hanya saja lompatan pengecekan pada proses pencarian *string* yang membedakan cara kerja antara kedua algoritma ini. Pada algoritma *Boyer-Moore* proses pencocokan setiap karakter diawali dari sisi kanan ke sisi kiri, sedangkan *Knuth-Morris-Pratt* melakukan pencocokan mulai dari kiri ke kanan.[4]

Berikut ini merupakan tahapan dari algoritma *Boyer-Moore* pada saat mencocokkan *string*:

1. Algoritma ini memulai pencocokkan *pattern* pada awal teks.
2. Dari sisi kanan ke sisi kiri, algoritma ini melakukan pencocokan setiap karakter yang berada di *pattern* dengan karakter pada teks yang sesuai, hingga salah satu dari kondisi berikut ini terpenuhi:
 - a. Karakter pada teks tidak cocok jika dibandingkan dengan karakter *pattern*.
 - b. Semua karakter *pattern* terjadi kecocokan dengan karakter pada teks. Selanjutnya algoritma akan memberitahu posisi penemuan karakter.
3. Algoritma melakukan penggeseran *pattern* dengan nilai penggeseran *bad-character* (OH) dan penggeseran *good-suffix* (MH), lalu mengulangi langkah ke-2 hingga *pattern* sampai di ujung teks.[9]

Berikut ini contoh penerapan algoritma *Boyer-Moore* dalam proses pencarian *pattern* berupa tanda baca pada teks:

Teks (S) = “sistem.”

Pattern (P) = “.” (tanda baca titik)

Tahap pencarian *pattern* (p) pada teks (S) yang ditunjukkan pada Gbr. 5

Pattern	Karakter lain	“.” (tanda baca titik)
OH	1	0

Gbr. 5 Penyelesaian Nilai OH Pada Tanda Baca

Langkah 1 penyelesaian penggeseran *pattern* ditunjukkan pada Gbr. 6

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern	.						

Gbr. 6 Langkah 1 Pergeseran Pattern Tanda Baca

Karakter “.” pada *pattern* tidak ditemukan kecocokan jika dibandingkan dengan karakter “s” di dalam teks. Maka selanjutnya akan terjadi penggeseran karakter sebanyak 1 loncatan sesuai dengan nilai tabel OH pada *pattern* karakter lain, dimana penggeseran tersebut ditunjukkan pada Gbr. 7

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern		.					

Gbr. 7 Langkah 2 Pergeseran Pattern

Pada Gbr. 7 menunjukkan karakter “.” di dalam *pattern* masih terjadi ketidakcocokan dengan karakter “i” yang ada dalam teks. Selanjutnya penggeseran dilakukan kembali hingga karakter *pattern* terjadi kecocokan seperti yang ditunjukkan pada Gbr. 8

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern			.				

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern				.			

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern					.		

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern						.	

Posisi teks	1	2	3	4	5	6	7
Teks	s	i	s	t	e	m	.
Pattern							.

Gbr. 8 Langkah Pergeseran Pattern Tanda Baca Hingga Akhir

Sedangkan untuk contoh penerapan algoritma *Boyer-Moore* dalam proses pencarian *pattern* berupa kata yang harus menggunakan huruf kapital antara lain:

Teks (S) = “Hari Rabu”

Pattern (P) = “Rabu”

Tahap penyelesaian nilai OH pada *pattern* ditunjukkan pada Gbr. 9

Pattern	Karakter Lain	R	a	b	u
OH	4	3	2	1	0

Gbr. 9 Penyelesaian nilai OH Huruf kapital

Langkah 1 pergeseran pattern huruf kapital dapat dilihat pada Gbr. 10

Posisi teks	1	2	3	4	5	6	7	8	9
Teks	H	a	r	i		R	a	b	u
Pattern	R	a	b	u					

Gbr. 10 Langkah 1 Pergeseran Pattern Huruf kapital

Pada Gbr 10 karakter “u” pada *pattern* dan karakter “i” pada teks tidak ditemukan kecocokan. Maka terjadi pergeseran karakter sebanyak 4, dimana pergeseran tersebut ditunjukkan pada Gbr. 11

Posisi teks	1	2	3	4	5	6	7	8	9
Teks	H	a	r	i		R	a	b	u
Pattern					R	a	b	u	

Gbr. 11 Langkah 2 Pergeseran Pattern Huruf Kapital

Pergeseran pada Gbr. 11 menampilkan karakter “u” di dalam *pattern* masih terjadi ketidakcocokan dengan karakter “b” yang ada dalam teks. Selanjutnya pergeseran dilakukan kembali hingga terjadi kecocokan pada setiap karakter. Seperti yang ditunjukkan pada Gbr. 12

Posisi teks	1	2	3	4	5	6	7	8	9
Teks	H	a	r	i		R	a	b	u
Pattern						R	a	b	u

Gbr. 12 Pergeseran Akhir Pattern Huruf Kapital

E. Presisi, Recall, dan Akurasi

Pengukuran kinerja pada sistem ini menggunakan *confusion matrix* yang merupakan alat evaluasi visual yang digunakan dalam pembelajaran mesin. Kolom yang terdapat pada *Confusion Matrix* mewakili hasil nilai prediksi, sedangkan baris mewakili nilai yang sebenarnya.[10] Terdapat 4 (empat) nilai klasifikasi, yaitu:

1. TP singkatan dari *True Positive*, merupakan data positif yang terdeteksi benar oleh sistem.
 2. FP singkatan dari *False Positive*, merupakan data positif tetapi terdeteksi salah oleh sistem
 3. TN singkatan dari *True Negative*, merupakan data negatif yang terdeteksi benar oleh sistem.
 4. FN singkatan dari *False Negative*, merupakan data negatif yang terdeteksi salah oleh sistem.
- confusion matrix* dapat dilihat pada Tabel 2.

TABEL 2
CONFUSION MATRIX

Nilai prediksi	Nilai Sebenarnya	
	Positif	Negatif
Positif	TP (True Positive)	FP (False Positive)
Negatif	FN (False Negative)	TN (True Negative)

Untuk lebih jelasnya berikut ini contoh hasil prediksi kesalahan kata dalam penggunaan huruf kapital yang ditunjukkan pada Tabel 3.

TABEL 3
CONTOH HASIL PREDIKSI KESALAHAN HURUF KAPITAL

Kata	Status Sebenarnya	Hasil Prediksi
Indonesia	Penulisan huruf kapital benar	Penulisan huruf kapital benar
Malaysia	Penulisan huruf kapital benar	Penulisan huruf kapital salah
senin	Penulisan huruf kapital salah	Penulisan huruf kapital benar
agustus	Penulisan huruf kapital salah	Penulisan huruf kapital salah
Lineke	Penulisan huruf kapital benar	Penulisan huruf kapital benar

Dari hasil prediksi pada tabel diatas, ada 4 kasus yang terjadi, yaitu:

1. True Positive (TP): Kasus dimana kata diprediksi benar (penulisan huruf kapital benar), dan sebenarnya kata tersebut memang benar penulisan huruf kapitalnya. Dalam contoh di atas adalah kata Indonesia dan Lineke.
2. True Negative (TN): Kasus dimana kata diprediksi salah (penulisan huruf kapital salah), dan sebenarnya kata tersebut memang salah dalam penulisan huruf kapitalnya. Dalam contoh di atas adalah kata Agustus.
3. False Positive (FP): Kasus dimana kata diprediksi benar (penulisan huruf kapital benar), ternyata kata tersebut salah dalam penulisan huruf kapitalnya. Dalam contoh di atas adalah kata Malaysia.
4. False Negative (FN): Kasus dimana kata diprediksi salah (penulisan huruf kapital salah), tetapi ternyata kata tersebut benar dalam penulisan huruf kapitalnya. Dalam contoh di atas adalah kata Senin.

Nilai presisi, *recall* dan akurasi akan diperoleh berdasarkan nilai klasifikasi *confusion matrix*. Nilai Presisi adalah banyaknya peluang informasi yang benar dari keseluruhan informasi yang didapat dari sistem. Nilai *recall* merupakan banyaknya peluang informasi benar yang didapat kembali dari sistem dibandingkan dengan informasi yang benar. Sementara itu, nilai akurasi merupakan gambaran keakuratan sistem untuk dapat mengklasifikasi data dengan benar. Rumus presisi berdasarkan tabel *confusion matrix* didapat:

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (1)$$

Berdasarkan tabel *confusion matrix*, untuk perhitungan *recall* menggunakan rumus:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

Sedangkan untuk perhitungan akurasi berdasarkan tabel *confusion matrix* menggunakan rumus:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

III. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

Sistem kesalahan tanda baca dan huruf kapital digunakan sebagai perangkat untuk membantu mencari kesalahan yang terdapat di karya tulis ilmiah. Berikut adalah tampilan dari sistem deteksi kesalahan tanda baca dan huruf kapital :

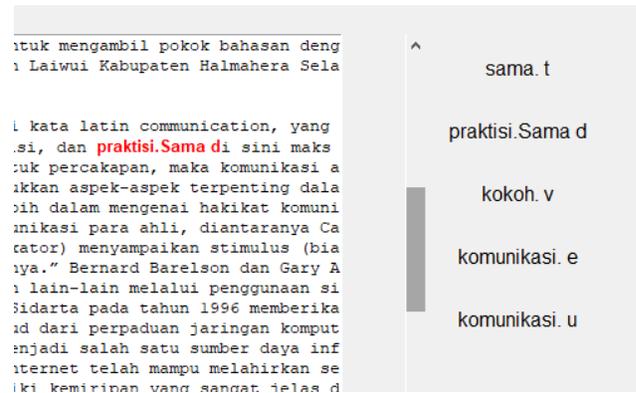


Gbr. 13 Tampilan sistem deteksi

Pada Gbr. 13 menunjukkan tampilan sistem deteksi. Halaman ini terdapat tombol *open* untuk membuka dokumen yang akan dideteksi. Nama dari dokumen tersebut akan tampil pada tampilan. Setelah membuka dokumen, isi teks pada dokumen akan ditampilkan pada kolom teks. Pada halaman ini terdapat tombol deteksi yang berguna untuk memulai proses deteksi. Hasil deteksi kesalahan akan ditandai dengan tulisan berwarna merah dan kata atau tanda baca yang terdeteksi salah akan ditampilkan di kolom kesalahan.

B. Pengujian Algoritma Boyer-Moore

Pengujian sistem dilakukan dengan melakukan perbandingan antara hasil deteksi analisa manual dengan hasil deteksi sistem. Data pengujian akan dibaca terlebih dahulu oleh reviewer, lalu reviewer akan mencari kesalahan dalam penggunaan tanda baca dan penulisan huruf kapital pada data uji tersebut. Pengujian yang dilakukan oleh sistem dengan hasil keluaran berupa identifikasi kesalahan yang ditandai dengan warna merah. Kemudian hasil dari deteksi secara manual dan secara sistem akan dibandingkan. Hasil *output* dari sistem pada data uji 1 terdapat pada Gbr. 14



Gbr. 14 *output* dari sistem

Sebanyak 5 data uji yang diambil acak digunakan untuk proses pengujian. Adapun Informasi dari data uji 1 terdapat pada Tabel 4.

TABEL 4
INFORMASI DATA Uji 1

Nomor	1
Nama dokumen	Jurnal 1 - Manfaat Internet Sebagai Media Komunikasi Bagi Remaja.docx
Judul karya ilmiah	Manfaat Internet Sebagai Media Komunikasi Bagi Remaja Di Desa Air Mangga Kecamatan Laiwui Kabupaten Halmahera Selatan
Jumlah tanda baca dan kata yang seharusnya menggunakan huruf kapital	362

Berikut ini contoh penerapan algoritma *Boyer-Moore* dalam melakukan pencarian *pattern* yang merupakan kata pada korpus nama dengan teks yang terdapat di data uji 1:

Teks (S) = bukunya Riswand

Pattern (P) = Riswand

Proses pencarian *pattern* (P) pada teks (S) yang ditunjukkan pada Gbr. 15

Pattern	R	i	s	w	a	n	d
OH	6	5	4	3	2	1	0

Gbr. 15 Penyelesaian nilai OH contoh data uji 1

Langkah 1 penyelesaian pergeseran *pattern* ditunjukkan pada Gbr. 16

Posisi (P)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Teks	b	u	k	u	n	y	a		R	i	s	w	a	n	d
Pattern	R	i	s	w	a	n	d								

Gbr. 16 Langkah 1 pergeseran *pattern* contoh data uji 1

Pada Gbr 16 karakter “d” pada pattern tidak ditemukan kecocokan dengan karakter “a” pada teks. Maka terjadi pergeseran karakter sebanyak 2 jika dilihat dari tabel OH pada pattern “a”, dimana pergeseran tersebut ditunjukkan pada Gbr. 17

Posisi (P)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Teks	b	u	k	u	n	y	a		R	i	s	w	a	n	d
Pattern									R	i	s	w	a	n	d

Gbr. 17 Akhir pergeseran pattern contoh data uji 1

Pada Gbr. 17 menampilkan bahwa kata “Riswandi” dalam *pattern*, cocok dengan kata “Riswandi” yang ada pada teks. Hal tersebut menyatakan bahwa kata yang ada pada korpus nama terdapat pada teks data uji 1. Selanjutnya kata atau karakter yang ditemukan akan diidentifikasi kebenarannya.

Hasil proses identifikasi akan ditulis keterangan yang bernilai *true positive*, *true negative*, *false negative* dan *false positive*. Dari identifikasi tersebut kemudian akan didapatkan perhitungan akurasi, presisi, dan *recall*. Pada Tabel 5 terdapat hasil identifikasi dan perbandingan daftar kesalahan pada data uji 1.

TABEL 5
HASIL IDENTIFIKASI DAN PERBANDINGAN DAFTAR KESALAHAN PADA DATA UJI 1

Nilai prediksi	Nilai Sebenarnya	
	Positif	Negatif
Positif	347	9
Negatif	1	6

Sedangkan untuk daftar hasil hitung rata-rata nilai akurasi, presisi, dan *recall* dari sistem berdasarkan pada 5 data uji dapat dilihat pada Tabel 6.

TABEL 6
DAFTAR NILAI PRESISI, RECALL DAN AKURASI DARI HASIL PENGUJIAN

No.	Nama Data Uji	Presisi	Recall	Akurasi
1.	Jurnal 1 - Manfaat Internet Sebagai Media Komunikasi Bagi Remaja.docx	0.974	0.997	0.972
2.	Jurnal 2 - Pemanfaatan Media Sosial Untuk Efektifitas Komunikasi.docx	0.988	0.994	0.982
3.	Jurnal 3 revisi - Teknologi Pendidikan Sebagai Pembelajaran Kompetitif Untuk	0.981	0.996	0.961

	Meningkatkan Prestasi Siswa.docx			
4.	Mazidhatul Ilmiah-Esai fix.docx	0.944	0.947	0.832
5.	PANCASILA SEBAGAI SISTEM FILSAFAT.docx	0.958	0.949	0.84

Setelah didapatkan hasil pengujian terhadap 5 data uji, ketidak akuratan masih ditunjukkan oleh sistem sehingga nilai presisi yang didapat kurang dari 1. Meskipun sistem berhasil mendeteksi hampir seluruh kesalahan tanda baca dan huruf kapital pada data uji, namun jika dilihat dari perbandingan antara hasil pemeriksaan manual dan pemeriksaan menggunakan sistem, masih terdapat beberapa kalimat yang benar penggunaan tanda bacanya apabila dilakukan pemeriksaan secara manual, namun pada pemeriksaan sistem kalimat tersebut termasuk kesalahan.

Seperti pada kata “teguh” yang pada penelitian ini termasuk dalam korpus nama orang, namun dalam sistem kata tersebut termasuk kesalahan karena tidak menggunakan huruf kapital di awal kata. Kata “teguh” yang digunakan pada karya ilmiah ini mengandung arti lain yang bukan nama. Contoh lain seperti pada data uji 5, terdapat kesalahan tanda baca koma (.). Jika pada aturan bahwa setiap tanda baca koma diikuti huruf kecil, akan tetapi karena setelah tanda baca koma diikuti oleh kata HAM yang merupakan singkatan, maka harus menggunakan huruf kapital dalam penulisannya. Kelemahan seperti ini akan mempengaruhi hasil dari nilai presisi yang diperoleh. Kesalahan seperti itu banyak didapati pada dokumen uji lain yang telah dideteksi oleh sistem.

IV. KESIMPULAN

Berdasarkan hasil yang diperoleh dalam penelitian, diambil kesimpulan bahwa sistem berhasil mendeteksi kesalahan penggunaan tanda baca dan penulisan huruf kapital pada karya tulis ilmiah menggunakan algoritma *Boyer-Moore*. Sistem menghasilkan nilai rata-rata presisi dan *recall* sebesar 0,969 dan 0,976. Sedangkan untuk nilai rata-rata tingkat akurasi pada sistem nilainya sebesar 0,917. Masih ada beberapa kesalahan yang terjadi pada sistem, untuk mengurangi kesalahan deteksi tanda baca yang terjadi karena adanya singkatan, bisa ditambahkan korpus berisi singkatan kata dan juga penambahan aturan baru. Lalu untuk menambahkan nilai akurasi pada sistem, dan menghindari kesalahan dalam mendapatkan informasi dari suatu kata, perlu ditambahkan metode pengolahan bahasa alami agar informasi yang didapatkan dari setiap kata pada karya tulis ilmiah lebih akurat dan aturan tidak terbatas pada penggunaan huruf kapital dan pemeriksaan spasi saja.

UCAPAN TERIMA KASIH

Puji dan rasa syukur senantiasa penulis panjatkan kepada Allah SWT karena telah diberi kemudahan serta kelancaran

dalam proses penyelesaian penelitian ini. Serta penulis ucapkan terimakasih kepada keluarga, teman-teman, serta pihak terkait lainnya atas dukungan dan juga bimbingan sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

- [1] S. Andayani, "Peningkatan Kemampuan Menulis Kembali Berita yang Dibacakan ke Dalam Beberapa Kalimat Melalui Teknik Pelatihan Terbimbing pada Siswa Kelas VII Semester Ganjil SMP Karya Bhakti Gading Rejo Kabupaten Pringsewu Tahun Pelajaran 2011/2012," *Digit. Repos. Unila*, p. 7, 2012.
- [2] F. Kurniadi, "PENULISAN KARYA TULIS ILMIAH MAHASISWA DENGAN MEDIA APLIKASI PENGOLAH KATA," *AKSIS J. Pendidik. Bhs. dan Sastra Indones.*, 2017, doi: 10.21009/aksis.010208.
- [3] R. N. E. Anggraini, M. A. Zinni, and S. Rochimah, "KAKAS BANTU PENDETEKSI KESALAHAN TANDA BACA PADA KARYA TULIS ILMIAH," *JUTI J. Ilm. Teknol. Inf.*, 2016, doi: 10.12962/j24068535.v14i1.a517.
- [4] F. T. Waruwu and R. Mandala, "Perbandingan algoritma knuth morris pratt dan boyer moore dalam pencocokan string pada aplikasi kamus bahasa nias," *J. Ilm. INFOTEK*, 2017.
- [5] R. I. Borman, "Penerapan String Matching Dengan Algoritma Boyer Moore Pada Aplikasi Font Italic Untuk Deteksi Kata Asing," *J. Teknoinfo*, 2016, doi: 10.33365/jti.v10i2.9.
- [6] R. Rahim, A. S. Ahmar, A. P. Ardyanti, and D. Nofriansyah, "Visual Approach of Searching Process using Boyer-Moore Algorithm," in *Journal of Physics: Conference Series*, 2017, doi: 10.1088/1742-6596/930/1/012001.
- [7] T. F. ASTUTI, "Penggunaan Ejaan Bahasa Indonesia Pada Karya Ilmiah Kelas Xi Ipa Smas Immanuel Bandarlampung Tahun 2011/2012," *Digit. Repos. Unila*, 2012, doi: 10.1017/CBO9781107415324.004.
- [8] E. Rahmanita, "PENCARIAN STRING MENGGUNAKAN ALGORITMA BOYER MOORE PADA DOKUMEN," 2014.
- [9] Tomy Satria Alasi, "Algoritma Boyer Moore Untuk Penyaringan Pesan Teks Menggunakan Perbandingan Kata Yang Sama," *Semin. Nas. Sains Teknol. Inf.*, 2018.
- [10] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny)*, 2020, doi: 10.1016/j.ins.2019.06.064.