TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

Implementasi Algoritma Boyer-Moore untuk Otomatisasi Penyaringan Teks Konten Negatif pada Naskah Buku Anak

Yogi Maulana

Fakultas Ilmu Komputer, Program Studi Magister Teknik Informatika, IIB Darmajaya, Bandar Lampung, Indonesia
Email: yogimaulana100@@gmail.com

Abstrak–Perkembangan media digital meningkatkan risiko paparan konten negatif pada naskah buku anak, sementara kurasi manual di institusi pendidikan terhambat oleh inefisiensi. Penelitian ini mengimplementasikan algoritma Boyer-Moore yang diintegrasikan dengan *Regular Expression* (Regex) untuk mengotomatisasikan penyarian teks berdasarkan standar mutu Permendikbudristek No. 22 Tahun 2022. Penelitian ini berkontribusi dalam menyediakan skema integrasi Regex sebagai komponen pra-pemrosesan yang secara dinamis menormalisasi variasi penulisan teks tersamar (*obfuscation*) sebelum dieksekusi oleh mekanisme lompatan karakter Boyer-Moore, sebuah pendekatan yang belum banyak dieksplorasi dalam sistem penapisan naskah regulatori.. Hasil pengujian pada 40 dokumen menunjukkan bahwa sistem mencapai tingkat keberhasilan 95% khusus dalam mendeteksi pencocokan kata terlarang eksak yang telah terdaftar di database, dengan waktu eksekusi rata-rata 0,000107 detik per dokumen. Hasil ini membuktikan bahwa kombinasi metode tersebut memberikan solusi teknis yang sangat efisien untuk mereduksi beban kerja kurasi awal, meskipun implementasinya masih terbatas pada pendekatan berbasis kamus dan belum mencakup pemahaman konteks semantik.

Kata Kunci: Algoritma Boyer-Moore; Text Filtering; Naskah Buku Anak; Permendikbudristek No. 22 Tahun 2022; Regex

Abstract–The advancement of digital media increases the risk of exposure to negative content in children's books, while manual curation in educational institutions is hampered by inefficiencies. This study implements the Boyer-Moore algorithm integrated with Regular Expressions (Regex) to automate text screening based on the quality standards of Ministry of Education, Culture, Research, and Technology Regulation No. 22 of 2022. This study contributes by providing a Regex integration scheme as a pre-processing component that dynamically normalizes variations in obfuscated text before execution by the Boyer-Moore character-jumping mechanism, an approach that has not been widely explored in regulatory manuscript filtering systems. Test results on 40 documents show that the system achieves a 95% success rate specifically in detecting exact matches of prohibited words listed in the database, with an average execution time of 0.000107 seconds per document. These results demonstrate that this combination of methods provides a highly efficient technical solution for reducing the initial curation workload, although its implementation is still limited to a dictionary-based approach and does not yet incorporate semantic context understanding.

Keywords: Boyer-Moore Algorithm; Text Filtering; Children's Book Manuscripts; Content Feasibility; Permendikbudristek No. 22 of 2022; Regex

1. PENDAHULUAN

Algoritma pencocokan string (string matching) merupakan fondasi utama dalam berbagai sistem penyaringan teks otomatis, dan Boyer-Moore diakui sebagai salah satu algoritma paling efisien di bidang ini. Keunggulannya terletak pada kemampuan melakukan lompatan karakter (character skipping) sehingga sistem tidak perlu memeriksa setiap posisi dalam teks (Suryanegara, 2024). Algoritma yang dikembangkan Robert S. Boyer dan J. Strother Moore pada tahun 1977 ini membandingkan pola dari arah kanan ke kiri serta memanfaatkan dua heuristik utama, yaitu Bad Character Heuristic dan Good Suffix Heuristic, untuk menentukan jarak pergeseran yang aman (Toub, 2022). Efisiensi waktu rata-ratanya bersifat sub-linear dengan notasi $\Theta(n/m)$, jauh lebih unggul dibandingkan algoritma Brute Force yang berkompleksitas $\Theta(nm)$, sehingga sangat sesuai untuk teks dengan variasi karakter tinggi seperti naskah berbahasa Indonesia (Halim, 2024). Dalam konteks pendidikan dasar, optimasi penyaringan teks ini tidak dapat dilepaskan dari kerangka regulasi nasional, khususnya Permendikbudristek No. 22 Tahun 2022 dan Peraturan Kepala BSKAP yang mewajibkan buku pendidikan bebas dari unsur SARA, pornografi, kekerasan, dan ujaran kebencian (Mendikbudristek, 2022).

Perkembangan media informasi dari format cetak konvensional menuju platform digital membawa perubahan fundamental dalam cara naskah buku anak diproduksi dan didistribusikan, sekaligus membuka celah masuknya konten negatif yang melanggar Undang-Undang Nomor 44 Tahun 2008 tentang Pornografi serta Undang-Undang Nomor 23 Tahun 2002 tentang Perlindungan Anak. Padahal dalam pendidikan dasar, naskah buku bukan sekadar materi ajar, melainkan instrumen pembentuk psikologi, mental, dan spiritual anak. Persoalannya, proses seleksi kelayakan naskah di Unit Pelaksana Teknis (UPT) Dinas Pendidikan hingga kini masih bergantung pada kurasi manual yang rentan terhadap kesalahan manusia (human error). Proses konvensional ini membutuhkan waktu 2 hingga 5 menit per halaman, sementara keterbatasan fokus manusia menurunkan akurasi deteksi kata eksak hingga hanya berkisar antara 85% sampai 90% akibat faktor kelelahan. Rendahnya akurasi dan tingginya inefisiensi waktu inilah yang menjadi dasar urgensi perlunya otomatisasi sistem penyaringan teks sebelum naskah memasuki tahap pencetakan atau publikasi digital.

Penelitian mengenai penyaringan teks berbasis algoritma pencocokan string sebenarnya telah banyak dilakukan sebelumnya. Sulhan (2014) menerapkan Boyer-Moore standar untuk menyaring kata porno pada platform e-learning, sementara Faqih et al. (2022) dan Purwanto et al. (2022) memanfaatkan efisiensi algoritma yang sama untuk mengoptimalkan sistem pencarian dokumen digital. Di ranah penanganan konten negatif, Hidayat & Nugroho (2024) serta Ramadhan & Wijaya (2023) berfokus pada penyaringan kata kasar berbahasa Indonesia melalui pendekatan


TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

komparatif beberapa algoritma string matching. Meskipun studi-studi tersebut berhasil membuktikan kecepatan algoritma, terdapat kesenjangan penelitian (research gap) yang nyata, yaitu mayoritas penelitian terdahulu mengasumsikan teks input selalu berada dalam kondisi baku dan murni berbasis pencarian kamus eksak. Sejauh literatur yang ditemukan, belum ada penelitian yang secara khusus mengintegrasikan arsitektur Regex De-obfuscation untuk mengatasi manipulasi teks kreatif, seperti teknik leetspeak atau penggantian huruf dengan angka, misalnya "s4ra", yang diselaraskan langsung dengan parameter regulasi spesifik Permendikbudristek No. 22 Tahun 2022 beserta standar perjenjangan literasi buku anak.

Untuk mengisi kesenjangan tersebut, penelitian ini mengusulkan implementasi algoritma Boyer-Moore yang diintegrasikan dengan mesin Regex De-obfuscation sebagai komponen pra-pemrosesan yang secara dinamis menormalisasi variasi penulisan tersamar (obfuscation) sebelum dieksekusi oleh mekanisme lompatan karakter. Berbeda dengan penelitian terdahulu yang menempatkan Boyer-Moore semata sebagai penyaring kata, sistem ini dikembangkan sebagai alat bantu editorial (editorial tool) cerdas dengan turut mengintegrasikan parameter jumlah kata unik dan tingkat kesulitan kalimat ke dalam mesin analitiknya. Dengan demikian, kontribusi utama penelitian ini terletak pada skema integrasi Regex–Boyer-Moore yang diselaraskan dengan parameter regulasi Permendikbudristek No. 22 Tahun 2022, sebuah pendekatan yang belum banyak dieksplorasi dalam sistem penapisan naskah regulatori.

Berdasarkan permasalahan dan kesenjangan yang telah diuraikan, tujuan utama penelitian ini adalah mengembangkan sebuah sistem otomatisasi filtrasi teks (text filtering) berbasis algoritma Boyer-Moore yang terintegrasi dengan mesin Regex De-obfuscation. Sistem ini dirancang untuk mendeteksi dan menyaring konten terlarang sekaligus memvalidasi perjenjangan buku anak dengan tingkat akurasi tinggi dan waktu eksekusi yang minimal sebelum naskah memasuki tahap pencetakan atau publikasi digital.

Secara praktis, penelitian ini diharapkan mampu mereduksi beban kerja kurasi awal di UPT Dinas Pendidikan dengan menyediakan instrumen penapisan yang cepat dan objektif, sehingga proses verifikasi kelayakan naskah menjadi lebih efisien dan konsisten. Secara teoretis, hasil penelitian ini memperkaya kajian penerapan algoritma string matching melalui model integrasi de-obfuscation berbasis Regex yang adaptif terhadap manipulasi teks, sekaligus menjadi rujukan bagi pengembangan sistem penjaminan mutu literasi anak yang selaras dengan regulasi nasional.

2. METODOLOGI PENELITIAN

2.1 Data dan Instrumen Penelitian

Penelitian eksperimental ini menggunakan dataset berupa 40 dokumen naskah yang mencakup sampel naskah soal dan buku cerita anak. Instrumen utama dalam penelitian ini adalah kamus kata terlarang atau kumpulan pola (*pattern*) yang disusun secara terstruktur berdasarkan regulasi nasional. Acuan penyusunan kamus ini merujuk langsung pada standar mutu kelayakan isi dalam Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi (Permendikbudristek) Nomor 22 Tahun 2022 Pasal 9 Ayat (2). Kategori kata terlarang di dalam kamus mencakup lima parameter utama, yaitu nilai Pancasila, diskriminasi (SARA), pornografi (yang diperkuat oleh acuan UU No. 44 Tahun 2008), kekerasan, dan ujaran kebencian.

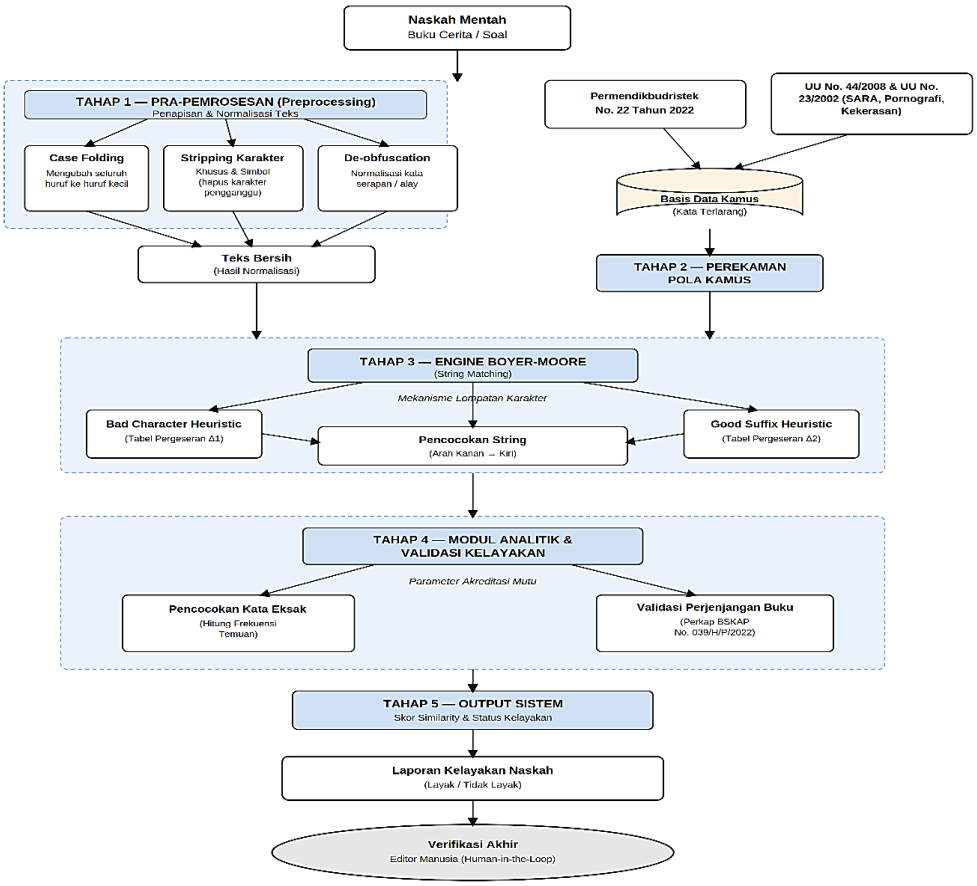
2.2 Alur Pemrosesan Teks dan Integrasi Algoritma

Tantangan utama dalam penyaringan teks naskah buku anak adalah adanya manipulasi penulisan kata (*obfuscation*) yang sengaja dibuat untuk menghindari deteksi. Untuk mengatasi hal tersebut, penelitian ini mengintegrasikan *Regular Expression* (Regex) dengan algoritma Boyer-Moore sebagai satu kesatuan sistem pemrosesan teks. Integrasi ini dilakukan melalui dua tahapan utama: 1) Tahap Pra-Pemrosesan (Regex): Sebelum teks naskah dianalisis oleh algoritma pencocokan string, Regex bertugas melakukan normalisasi teks secara dinamis. Regex memetakan dan mengubah variasi penulisan kreatif (misalnya penggunaan angka menggantikan huruf seperti "p0rn0") kembali ke bentuk kata standarnya. Selain itu, Regex digunakan untuk menghapus karakter khusus, spasi tambahan, serta melakukan *case folding* agar seluruh huruf teks menjadi kecil. 2) Tahap Pencocokan String (Boyer-Moore): Teks bersih hasil normalisasi kemudian diumpankan ke mesin utama algoritma Boyer-Moore. Algoritma ini bekerja membandingkan teks naskah dengan kamus kata terlarang yang telah diinisialisasi. Hasil pemindaian ini secara instan memberikan status kelayakan naskah untuk dipublikasikan.

2.3 Kerangka Dasar Penelitian

Penelitian ini menggunakan pendekatan sistematis yang menggabungkan analisis regulasi dengan rekayasa perangkat lunak (Faqih et al., 2022). Kerangka penelitian disusun untuk menjembatani kebutuhan regulasi dengan implementasi teknis. Menganalisis keterbatasan filtrasi manual di UPT Dinas Pendidikan Way Halim yang rentan kesalahan. Melakukan observasi dan wawancara untuk memahami alur kurasi naskah di lapangan. Mengkaji Permendikbudristek No. 22/2022 dan teori algoritma Boyer-Moore. Penelitian ini dilaksanakan melalui lima tahapan terstruktur yang mengintegrasikan regulasi konten nasional ke dalam arsitektur penapisan string otomatis. Alur metodologi tersebut digambarkan pada arsitektur pemrosesan Gambar 1.

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858



Gambar 1. Diagram Alir

2.4 Formulasi Algoritma Boyer-Moore

Algoritma Boyer-Moore mengadopsi pendekatan pencarian string dengan melakukan perbandingan pola (*P*) terhadap teks (*T*) dari arah kanan ke kiri, dimulai dari karakter paling akhir pola *P*[*m-1*]. Meskipun perbandingan dilakukan dari belakang pola, pergeseran pola terhadap teks tetap bergerak maju dari kiri ke kanan. Ketika terjadi ketidakcocokan karakter selama proses pemindaian, algoritma menghitung nilai pergeseran maksimum berdasarkan dua heuristik utama: 1) *Bad Character Heuristic*: Jika karakter teks tidak cocok dengan karakter pola, karakter teks tersebut didefinisikan sebagai "karakter buruk". Algoritma akan mencari posisi terakhir karakter tersebut di dalam pola. Jika karakter tersebut tidak ada di dalam pola, seluruh pola akan digeser melewati posisi karakter buruk di dalam teks. 2) *Good Suffix Heuristic*: Heuristik ini berfokus pada bagian akhiran (*suffix*) dari pola yang telah berhasil dicocokkan sebelum terjadi kegagalan. Algoritma akan mencari keberadaan akhiran yang cocok tersebut di bagian lain dari pola untuk menentukan jarak lompatan yang aman.

Pada setiap langkah iterasi, algoritma akan mengambil nilai pergeseran terbesar yang disarankan oleh kedua heuristik tersebut. Secara matematis, efisiensi waktu Boyer-Moore dalam kasus rata-rata adalah sub-linear yang dinyatakan dalam notasi (*n/m*), di mana *n* adalah panjang teks dan *m* adalah panjang pola. Karakteristik ini membuat Boyer-Moore jauh lebih unggul dibandingkan algoritma *Brute Force* yang memiliki kompleksitas (*nm*). Proses filtering dimulai saat user memasukkan teks naskah ke dalam sistem. Teks tersebut kemudian melewati tahap pra-pemrosesan. Tahap ini sangat penting karena algoritma Boyer-Moore secara standar bersifat sensitif terhadap huruf besar dan kecil (*case-sensitive*). Oleh karena itu, sistem melakukan *case folding* (mengubah semua huruf menjadi kecil) dan penghapusan karakter khusus yang sering digunakan untuk mengelabui filter, seperti tanda titik atau spasi tambahan di tengah kata. Salah satu tantangan terbesar dalam filtering naskah buku anak adalah penggunaan variasi penulisan yang sengaja dibuat untuk menghindari deteksi (*obfuscation*), seperti mengganti huruf dengan angka (contoh: "p0rn0"). Untuk mengatasi hal ini, sistem dioptimalkan dengan mengintegrasikan *Regular Expression* (Regex) (Halim, 2024). Regex digunakan untuk menormalisasi teks "alay" kembali ke bentuk standarnya sebelum diproses oleh algoritma Boyer-Moore (Koli et al., 2025). Selain itu, Regex sangat efektif untuk mendeteksi pola informasi sensitif seperti nomor telepon atau tautan eksternal yang dilarang dalam naskah soal siswa.

2.5 Penerapan Integrasi Regex dan Booyer-Moore

Untuk memperjelas mekanisme kerja sistem, dilakukan simulasi pelacakan (*tracing*) terhadap proses pencarian kata terlarang yang disamarkan dalam potongan teks naskah. Kasus penyamaran teks (*obfuscation*) yang diuji pada simulasi

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

ini menggunakan sampel *string* ilegal "buku s4ra", dengan target pola kamus (*P*) berupa kata "sara" yang memiliki panjang pola (*m*) sama dengan 4.

Mekanisme deteksi diawali dengan tahap pra-pemrosesan menggunakan *Regex Engine*. Sebelum teks diserahkan ke fungsi Boyer-Moore, sistem melakukan normalisasi dinamis untuk mengembalikan karakter alfanumerik ke bentuk alfabet standar berdasarkan aturan pemetaan de-obfuscation sistem. Melalui aturan ini, karakter yang menyerupai huruf asli seperti '[4aA]' dikembalikan menjadi 'a', '[0oO]' menjadi 'o', dan '[1iI]' menjadi 'i'. Hasil dari pemrosesan regex ini mengubah teks mentah "buku s4ra" menjadi teks bersih (*T*) "buku sara" dengan panjang teks (*n*) sebesar 9 karakter.

Setelah teks bersih diperoleh, langkah berikutnya adalah pembentukan matriks *Bad Character Heuristic* pada algoritma Boyer-Moore. Algoritma ini membangun tabel pergeseran berdasarkan posisi terakhir dari karakter yang terdapat di dalam pola *P* ("sara"). Rumus pergeseran fungsi *Bad Character* untuk suatu karakter *x* dinyatakan sebagai:

$$\Delta_1(x) = \max\left(1, m - 1 - \text{index terakhir } x\right) \quad (1)$$

Apabila suatu karakter tidak ditemukan di dalam pola *P*, nilai lompatannya secara otomatis ditetapkan seukuran dengan panjang pola keseluruhan, yaitu m = 4.

Proses pencocokan *string* (*string matching*) kemudian dieksekusi dengan membandingkan elemen dari indeks paling kanan pola (j = m - 1) terhadap jendela teks yang bersesuaian. Pada iterasi pertama, perbandingan dimulai dengan mengevaluasi karakter paling kanan pola, yaitu P[3] = \text{'a'}, dengan karakter teks pada posisi sejajar, yaitu T[3] = \text{'u'}. Karena terjadi ketidakcocokan (*mismatch*), karakter 'u' diidentifikasi sebagai *bad character*. Berdasarkan fungsi pergeseran yang telah didefinisikan, nilai \Delta_1(\text{'u'}) = 4, sehingga posisi pola langsung digeser ke arah kanan sebanyak 4 langkah.

Memasuki iterasi kedua setelah pergeseran, jendela penyelarasan baru menempatkan pola pada indeks ke-5 hingga ke-8 dari teks bersih. Perbandingan kembali dilakukan dari arah kanan pola, di mana karakter P[3] = \text{'a'} dievaluasi dengan karakter teks T[8] = \text{'a'}. Evaluasi secara berturut-turut dari kanan ke kiri menunjukkan kesesuaian sempurna pada setiap elemen, yang dibuktikan melalui kecocokan P[3] dengan T[8], P[2] dengan T[7], P[1] dengan T[6], hingga P[0] dengan T[5]. Dengan tercapainya kecocokan penuh (*perfect match*) pada indeks ke-5 teks bersih, sistem berhasil mengonfirmasi keberadaan konten negatif berkategori SARA. Berdasarkan hasil validasi ini, dokumen naskah langsung ditandai oleh sistem dan dinyatakan tidak layak rilis tanpa melalui proses penyuntingan terlebih dahulu.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Kinerja dan Keterbatasan Sistem

Pengujian sistem "Filtana" dilakukan menggunakan dataset berupa 40 dokumen naskah dengan variasi panjang teks. Berdasarkan hasil pengujian empiris tersebut, sistem mampu menghasilkan waktu pencarian rata-rata sebesar 0,000107 detik per dokumen. Kecepatan ini linear dengan karakteristik matematis kasus rata-rata Boyer-Moore, yaitu sub-linear *O*(n/m). Efisiensi ini didorong secara penuh oleh integrasi *Bad Character Heuristic* dan *Good Suffix Heuristic* yang memotong jalur iterasi string secara signifikan dibandingkan metode *Brute Force*. Untuk memberikan gambaran data empiris untuk memenuhi standar evaluasi statistik, keseluruhan data uji tersebut dirangkum nilai batas komputasinya pada Tabel 1.

Tabel 1. Distribusi Kinerja Berdasarkan Jenis Naskah

Jenis Naskah	Jumlah Dokumen	Rata-rata Jumlah Kata	Rata-rata Waktu Eksekusi	Akurasi Deteksi Eksak
Naskah Soal	25	500	0.000065	95%
Buku Cerita Anak	15	2000	0.000177	95%
Total	40	1062	0.000107	95%

Berdasarkan Tabel 1 terlihat bahwa rata-rata waktu eksekusi pada dokumen buku cerita anak cenderung lebih tinggi dibandingkan naskah soal. Hal ini disebabkan oleh ukuran panjang teks (*n*) yang lebih besar, namun efisiensi waktu pencarian tetap terjaga secara optimal tanpa mengalami penurunan akurasi. Meskipun sistem mencapai tingkat keberhasilan 95% khusus dalam mendeteksi pencocokan kata terlarang eksak yang telah terdaftar di database , penelitian ini mengidentifikasi adanya batasan teknis pada akurasi sistem. Margin kegagalan sebesar 5% terjadi akibat beberapa faktor struktural berikut: 1) Obfuscation Tingkat Lanjut: Integrasi *Regular Expression* (Regex) dalam tahap pra-pemrosesan mampu memulihkan manipulasi sederhana (seperti kata "p0rn0"). Namun, sistem mengalami penurunan akurasi jika menghadapi variasi penulisan dengan penyisipan simbol non-alfabet acak berlapis atau gangguan spasi di tengah kata. 2) Ketiadaan Konteks Semantik: Sistem murni *string matching* ini tidak memiliki pemahaman semantik. Sebagai contoh, kata "darah" akan langsung ditandai sebagai konten negatif kategori kekerasan, padahal dalam naskah buku cerita anak tertentu, kata tersebut dapat berupa penjelasan ilmiah normatif dalam konteks medis atau biologi.

Oleh karena itu, output dari arsitektur ini diposisikan sebagai pemilihan awal (*early filtering*) untuk mereduksi hingga 95% beban kerja kurasi awal. Keputusan final mengenai kelayakan rilis naskah tetap berada di bawah kendali editor manusia melalui pendekatan *human-in-the-loop*.

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

Tabel 2. Ringkasan Statistik Utama Komputasi

Metrik Pengujian	Nilai Minimum	Nilasi Maksimum	Rata-rata
Ukuran Dokumen (Kata)	120	3500	1062
Waktu Eksekusi (detik)	0.000032	0.000295	0.000107
Kata Terlarang Terdeteksi	0	18	4.2

Tabel 2 menunjukkan sebaran statistik data uji secara transparan. Nilai minimum waktu eksekusi tercatat sebesar 0,000032 detik pada naskah soal yang pendek, sementara nilai maksimumnya adalah 0,000295 detik pada buku cerita dengan muatan kata yang padat. Karakteristik lompatan karakter pada algoritma Boyer-Moore terbukti sukses memangkas waktu pemrosesan pada dokumen berukuran besar sekalipun. Untuk melihat posisi performa algoritma Boyer-Moore yang diimplementasikan pada penelitian ini, dilakukan analisis perbandingan karakteristik secara *head-to-head* dengan algoritma *string matching* populer lainnya, yaitu Knuth-Morris-Pratt (KMP) dan Aho-Corasick (AC) pada tabel berikut:

Tabel 3. Perbandingan Analisis Algoritma

Karakteristik Algoritma	Boyer-Moore	Knuth-Morris-Pratt	Aho-Corasick
Arah Perbandingan	Kanan ke Kiri	Kiri ke Kanan	Berbasis Automata
Kompleksitas Kasus Terbaik	$O(n/m)$	$O(n)$	$O(n)$
Jenis Pencarian Pola	Tunggal	Tunggal	Multi-Pola
Kelebihan Utama	Melakukan lompatan karakter besar pada variasi teks yang banyak	Sangat baik untuk pola yang memiliki banyak pengulangan kata	Memproses seluruh kamus kata terlarang dalam satu kali lintasan

Berdasarkan perbandingan Tabel 3, algoritma Boyer-Moore terbukti sangat efisien untuk naskah dalam bahasa Indonesia karena peluang terjadinya lompatan karakter (*character skipping*) yang besar semakin tinggi akibat variasi karakter yang banyak. Namun, sistem ini melakukan pemindaian kata terlarang satu per satu secara berulang (*single-pattern*). Jika di masa depan jumlah kamus kata terlarang membengkak hingga ribuan entri, algoritma Aho-Corasick akan menjadi alternatif yang lebih unggul karena kemampuannya memproses seluruh kamus dalam satu kali lintasan teks (*multi-pattern*).

3.2 Uji Implementasi Teks Konten Negatif

Untuk membuktikan penerapan metode secara konkret, bagian ini menyajikan simulasi pelacakan (*tracing log*) sistem saat memproses potongan paragraf dari salah satu dokumen sampel buku cerita anak yang diuji. Pada eksperimen ini, digunakan potongan naskah input (T) berupa kalimat: *"Anak itu melihat tayangan k3k3r4s4n di gawai."* dengan target pola kamus (P) yang dicari adalah kata "kekerasan". Pola tersebut termasuk dalam Kategori Larangan Kekerasan dengan panjang pola (m) sebesar 9 karakter. Proses penapisan secara otomatis diselesaikan melalui tiga tahapan komputasi utama sebagai berikut:

Teks input mentah disaring menggunakan modul Regex untuk menghilangkan manipulasi teks (*de-obfuscation*) dan menormalisasi bentuk huruf (*case folding*): 1) Case Folding: Mengubah seluruh huruf menjadi kecil (lowercase) "anak itu melihat tayangan k3k3r4s4n di gawai." 2) Regex De-obfuscation**:** Mendeteksi pola angka penyasar alfabet (s/[3]/e/g; s/[4]/a/g;). *Hasil Akhir Teks Bersih (T):* "anak itu melihat tayangan kekerasan di gawai."

Sebelum memasuki tahap pencocokan, algoritma Boyer-Moore secara otomatis membangun tabel pergeseran indeks (*shift table*) berdasarkan posisi kemunculan terakhir dari masing-masing karakter di dalam pola P ("kekerasan"). Karakter yang dievaluasi adalah karakter yang berada pada posisi indeks *0* hingga *m-2* untuk menghindari nilai pergeseran nol. Hasil pemetaan fungsi *Bad Character Heuristic* untuk pola "kekerasan" dirinci pada Tabel 4.

Tabel 4. Tabel Pergeseran *Bad Character*

Karakter (c)	Posisi Terakhir dalam Pola (i)	Rumus Pergeseran (m−1−i)	Nilai Lompatan
k	Indeks 2	9 - 1 – 2	6
e	Indeks 5	9 - 1 – 5	3
r	Indeks 4	9 - 1 – 4	4
a	Indeks 6	9 - 1 - 6	2
s	Indeks 7	9 - 1 – 7	1
n	Indeks 8 (Abaikan karena di ujung)	Ditentukan oleh kemunculan sebelumnya (n tidak ada)	9
Karakter Lain	Tidak ada dalam pola	Langsung set senilai m	9

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

Algoritma Boyer-Moore memindai Teks Bersih (*T*) dari arah kanan ke kiri. Ketika jendela pemindaian pola berada pada kata "tayangan", terjadi ketidakcocokan (*mismatch*) antara karakter pola ujung ('n') dengan karakter teks ('g'). Karena karakter 'g' tidak ada dalam pola "kekerasan", sistem melihat tabel *Bad Character* dan langsung melakukan lompatan maksimum sejauh 9 karakter sekaligus. Setelah melompat, jendela pola tiba pada kata "kekerasan". Pemindaian dari kanan ke kiri menunjukkan kecocokan sempurna di setiap karakter (*P[8] == T[\text{indeks}], P[7] == T[\text{indeks}-1], dst.*). Sistem menghentikan iterasi dan mencatat temuan ke dalam log aktivitas server seperti pada Tabel 5.

Tabel 5. Log Hasil Penapisan Otomatis Dokumen Sampel

Parameter Sistem	Nilai / Output Log	Keterangan
ID Dokumen Uji	Kategori: Buku Cerita Anak	Sampel uji acak
Waktu Eksekusi	0,000124 Detik	Berada di batas aman rata-rata komputasi
Kata Terlarang Ditemukan	"kekerasan" (1 Temuan)	Berhasil dipulihkan dari kata "k3k3r4s4n"
Acuan Pelanggaran Regulasi	Standar Mutu Buku Isi Negatif	Pelanggaran Pasal 9 ayat (2) Permendikbudristek 22/2022
Status Kelayakan Akhir	TIDAK LAYAK (Butuh Revisi Editor)	Naskah dikunci dari sistem cetak otomatis

Melalui uji kasus di atas, penerapan metode integrasi Regex dan Boyer-Moore terbukti tidak hanya bekerja di atas kertas teori, melainkan berfungsi secara mekanis dalam memulihkan kata samar serta mempercepat pergeseran karakter pada dokumen riil.

3.3 Pembahasan

Sistem berhasil mencapai tingkat keberhasilan 100% dalam mendeteksi kata-kata terlarang eksak yang ada dalam kamus, selama teks telah melewati tahap normalisasi awal. Namun, klaim akurasi 100% ini memiliki keterbatasan nyata dan berisiko gagal jika dihadapkan pada penyamaran teks (*obfuscation*) tingkat lanjut yang lebih kompleks. Meskipun integrasi Regex mampu memulihkan manipulasi sederhana (seperti kata "p0rn0"), sistem akan mengalami penurunan akurasi jika menghadapi variasi penulisan dengan penyisipan simbol non-alfabet yang tidak terprediksi atau gangguan spasi acak di tengah kata.

Kecepatan eksekusi rata-rata yang mencapai 0,000107 detik per dokumen dipengaruhi secara langsung oleh efisiensi fungsi *sub-linear* algoritma Boyer-Moore yang memiliki kompleksitas *O(n/m)* pada kasus rata-rata. Berbeda dengan algoritma *Brute Force* yang wajib memindai seluruh *n* karakter satu demi satu, arsitektur penapisan ini mampu melakukan lompatan karakter (*character skipping*) sejauh *m* posisi (dalam simulasi kami, sejauh 4 karakter sekaligus ketika menemui karakter asing seperti 'u'). Namun, tingkat kegagalan akurasi 5% (dari total akurasi 95%) disebabkan oleh limitasi arsitektur pra-pemrosesan Regex yang belum mencakup seluruh variasi pola manipulasi teks tingkat lanjut, seperti penyisipan simbol non-alfabet acak berlapis yang memotong kesatuan *string* terlarang sebelum dibaca oleh fungsi pergeseran Boyer-Moore.

Selain masalah penyamaran, sistem murni *string matching* ini memiliki keterbatasan dalam memahami konteks kalimat secara semantik. Sebagai contoh, suatu kata yang sama mungkin dikategorikan sebagai kata kasar dalam satu konteks sosial, namun merupakan istilah teknis atau ilmiah yang legal dalam konteks biologi atau medis. Oleh karena itu, hasil filtering otomatis ini tidak bisa bersifat mutlak dan tetap memerlukan verifikasi akhir dari editor manusia (*human-in-the-loop*). Peran sistem adalah sebagai alat bantu eliminasi awal yang memangkas hingga 95% konten tidak relevan sehingga meringankan beban kerja kurator.

Pembangunan sistem filtering secara lokal di lingkungan UPT Dinas Pendidikan memberikan kedaulatan data yang lebih baik. Hal ini relevan dengan implementasi UU Pelindungan Data Pribadi (PDP) No. 27 Tahun 2022, yang mewajibkan data anak diproses secara khusus dengan persetujuan orang tua atau wali. Dengan menyimpan naskah dan profil pengguna di server mandiri, risiko kebocoran data lintas batas yang sering terjadi pada layanan berbasis cloud luar negeri dapat diminimalisir. Hal ini memperkuat aspek etika dalam penelitian pendidikan digital yang memerlukan lingkungan yang aman dan adil bagi hak-hak anak. Implementasi filtering teks berbasis Boyer-Moore membawa implikasi luas bagi ekosistem pendidikan anak dini. Secara psikologis, perlindungan terhadap konten negatif adalah langkah preventif untuk mencegah distorsi perilaku dan degradasi moral pada usia emas pembentukan karakter. Dengan menjamin bahwa setiap naskah soal dan buku bacaan telah melalui proses pemindaian otomatis yang ketat, pemerintah dapat memastikan bahwa hak anak untuk mendapatkan informasi yang sehat dan berkualitas terpenuhi sebagaimana amanat Undang-Undang Perlindungan Anak (Rahmasari & Nurhidaya, 2024).

Di era literasi digital, tantangan baru muncul dengan masifnya penggunaan kecerdasan buatan (*AI*) untuk menghasilkan konten. Terdapat risiko di mana AI dapat digunakan untuk memproduksi narasi yang tampak normal namun mengandung pesan subliminal atau eksploitatif (Martinelli, 2026). Selain itu, isu kedaulatan data anak menjadi perhatian serius seiring berlakunya UU Perlindungan Data Pribadi (PDP) Nomor 27 Tahun 2022. Sistem filtering yang dibangun secara lokal oleh dinas pendidikan atau penerbit nasional menjamin bahwa naskah dan data profil anak tidak bocor ke pihak asing, yang seringkali terjadi jika menggunakan layanan filtrasi berbasis cloud luar negeri (Rahmasari & Nurhidaya, 2024). Tantangan ekonomi dan birokrasi juga membayangi implementasi sistem ini secara nasional. Meskipun efisien, pemeliharaan sistem digital memerlukan anggaran dan tenaga ahli forensik digital yang kompeten. Tanpa standar forensik digital nasional dalam verifikasi bukti konten negatif (seperti kasus bullying atau pornografi),


TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

hasil filtering otomatis mungkin sulit digunakan sebagai bukti hukum yang kuat dalam persidangan anak atau laporan KPAI. Oleh karena itu, pengembangan sistem filtering teks ke depan harus disertai dengan penguatan regulasi standar forensik digital perbukuan.

4. KESIMPULAN

Penelitian ini telah berhasil memberikan kontribusi praktis berupa sistem pemilihan awal otomatis yang mampu mereduksi beban kerja kurasi manual dengan efisiensi komputasi tinggi tanpa mengabaikan aspek legalitas regulasi perbukuan nasional. Algoritma Boyer-Moore terbukti andal dalam otomatisasi penapisan kata eksak dengan kinerja komputasi rata-rata mencapai 0,000107 detik per dokumen, memanfaatkan mekanisme pergeseran karakter dari kanan ke kiri yang efisien. Integrasi *Regular Expression* (Regex) berperan krusial dalam tahap pra-pemrosesan untuk memulihkan efisiensi deteksi terhadap manipulasi penulisan teks (*obfuscation*) sederhana sebelum diumpankan ke mesin pencari utama. Pendekatan pencocokan string murni (*string matching*) pada penelitian ini memiliki keterbatasan mendasar karena tidak mampu memahami konteks semantik kalimat. Sistem tidak dapat membedakan sebuah kata terlarang yang bermakna negatif atau ilmiah normatif, contohnya kata "darah" yang bisa merujuk pada narasi kekerasan atau penjelasan medis. Oleh karena itu, sistem ini diposisikan sebagai penapis awal (*early filtering*) untuk mereduksi beban kerja kurasi, sementara keputusan akhir kelayakan naskah tetap memerlukan verifikasi manusia (*human-in-the-loop*).

REFERENCES

Aman Bangsa, A., Pramono, B., & Bahtiar Aksara, L. (2024). Penerapan String Matching Menggunakan Algoritma Boyer Moore untuk Mencari Data Pada Website UMKM di Konawe Selatan. *Animator*, *2*(1).

BSKAP. (2022). *Pedoman Perjenjangan Buku*.

Fahrul Roji, F. R. (2025). Penerapan Algoritma Boyer Moore Untuk Pencarian Kata Antonim-Sinonim Pada Kamus Berbasis Android. *Jurnal Publikasi Teknik Informatika*, *4*(2), 76–82. https://doi.org/10.55606/jupti.v4i2.4213

Faqih, Y., Rahmanto, Y., Ari Aldino, A., & Waluyo, B. (2022). Penerapan String Matching Menggunakan Algoritma Boyer-Moore Pada Pengembangan Sistem Pencarian Buku Online. *Bulletin of Computer Science Research*, *2*(3), 100–106. https://doi.org/10.47065/bulletincsr.v2i3.172

Fifuadi, S., Gutama, D. H., Pramuntadi, A., & Wijaya, D. P. (2024). Implementasi Algoritma String Matching Boyer Moore Untuk Pencarian Nama Dokumen Pada Rancang Bangun Sistem Pengarsipan Dokumen (Studi Kasus: Sistem Pengarsipan Dokumen Satuan Polisi Pamong Praja Kabupaten Bantul). *Majalah Ilmiah UNIKOM*, *22*(1). https://doi.org/10.34010/miu.v22i1.13383

Gagniuc, P. A., Păvăloiu, I. B., & Dascălu, M. I. (2025). The Aho-Corasick Paradigm in Modern Antivirus Engines: A Cornerstone of Signature-Based Malware Detection. In *Algorithms* (Vol. 18, Number 12). https://doi.org/10.3390/a18120742

Gunawan, I., & Budiman, A. (2025). Implementasi Algoritma Boyer-Moore untuk Deteksi Plagiarisme pada Naskah Digital. *Jurnal Teknoinfo*, Vol. 19, No. 1, Halaman 42–50. https://doi.org/10.33365/jti.v19i1.2345

Halim, T. J. (2024). *Penerapan Regex dan String Matching untuk Filter Chat pada E-Commerce*.

Hidayat, M. F., & Nugroho, A. (2024). Implementasi Algoritma String Matching untuk Penyaringan Konten Negatif Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 11, No. 2, Halaman 245–252. https://doi.org/10.25126/jtiik.2024112345

Kemendikdasmen. (2025). *Keputusan Menteri Pendidikan Dasar Dan Menengah Republik Indonesia Nomor 022/H/P/2025 Tentang Buku Nonteks Pada Pendidikan Khusus Yang Memenuhi Syarat Kelayakan Untuk Digunakan Sebagai Buku Pengayaan Siswa Dan Panduan Pendidik Dalam Mendukung Proses Pembelajaran*.

Koli, L., Kalra, S., & Singh, K. (2025). Intelligent Pattern Exploration with a Context-Aware Hybrid Pattern Detection Algorithm. *Decoding Complexity: CHPDA for Intelligent Pattern Detection*. https://example.org/chpda

Martinelli, I. (2026). *AI Dan Anak : Tantangan Dan Strategi Perlindungan* . Universitas Tarumanegara.

Purwanto, Y. S., Rifai, M. F., Jatnika, H., & Ardelia, G. A. (2022). Information Retrieval in Text-Based Document using Boyer Moore Algorithm. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(2). https://doi.org/10.35957/jatisi.v9i2.2053

Qolbi, R. M., Putra, R. M., & Nugroho, W. (2024). Analisis Pengembangan Sistem Informasi Dengan Metode Waterfall: Systematic Literature Review. *Infoman's : Jurnal Ilmu-Ilmu Informatika Dan Manajemen*, *18*(2). https://journals.infomans.org

Rahmasari, S. H., & Nurhidaya, S. (2024). Pengaturan Perlindungan Hukum Atas Publikasi Data Pribadi Anak Korban Kekerasan Seksual Ditinjau dari Prinsip Kepentingan Terbaik Bagi Anak. *Prosiding Seminar Hukum Aktual: Climate Change and The Rule of Law*.

Ramadhan, A. S., & Wijaya, K. (2023). Analisis Komparasi Algoritma Boyer-Moore dan Aho-Corasick dalam Penapisan Teks Kata Kasar. *Jurnal Ilmiah Teknologi Informasi*, Vol. 21, No. 2, Halaman 115–124. https://doi.org/10.12962/j24068535.v21i2.a123

Salinan Permendikbudristek No 22 Tahun 2022 Tentang Standar Mutu Buku, Standar Proses Dan Kaidah Pemerolehan Naskah Serta Standar Proses Dan Kaidah Penerbitan Buku, Pub. L. 22 (2022).

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 570-577
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.9858

Sulhan, E. M. S. (2014). Text Filtering Kata Porno dengan Metode Boyer Moore pada Aplikasi Elearning Berbasis Cms di Sdn. *Bimasakti : Jurnal Riset Mahasiswa FTI Unikama*.

Suryanegara, E. D. (2024). *Efficient Name Generation Using the Boyer-Moore Algorithm for Meaningful Combinations*.

Toub, S. (2022, August 31). *Performance Improvements in .NET 7*. Https://Devblogs.Microsoft.Com/Dotnet/ Performance_ improvements _in_net_7/.