

ANALISIS STRING MATCHING METHOD PADA PENGEMBANGAN APLIKASI E-SUBMISSION V2.0 MENGGUNAKAN ALGORITMA BOYER MOORE

Ariyo Atmojo, Roni Habibi

Program Studi Teknik Informatika D4, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia
ariyoatmojo92@gmail.com, roni.habibi@ulbi.ac.id

ABSTRAK

Setiap saat untuk mengambil informasi yang dibutuhkan berdasarkan kata kunci yang diinginkan, setiap teks atau string yang terdapat dalam penyimpanan data (database) akan diurutkan, dipisahkan, dan diorganisir kembali. Akibatnya, penerapan algoritma Boyer-Moore dalam masalah pencocokan string dieksplorasi dalam penelitian ini. Ketika menganalisis teks dari kumpulan data, pencocokan string adalah masalah yang krusial. Dibandingkan dengan teknik pencocokan string lainnya, metode Boyer-Moore dikenal baik karena efisiensinya dalam mengidentifikasi pola-pola dalam teks berdasarkan kata kunci dengan memerlukan waktu yang lebih sedikit. Khususnya, sebagai mesin pencari, artikel ini berfokus pada pembahasan fase-fase kunci dalam proses pencocokan string dan ringkasan ide-ide kunci dari algoritma Boyer-Moore. Ide dasar di balik metode ini adalah mencocokkan pola dari kanan ke kiri daripada dari kiri ke kanan. Hal ini memungkinkan untuk menghindari beberapa karakter teks dan mengurangi jumlah perbandingan yang diperlukan. Penelitian ini menunjukkan nilai dari algoritma Boyer-Moore sebagai alat pencocokan string yang kuat dalam konteks penggunaan mesin pencari, membuka jalan bagi penelitian masa depan untuk memperluas aplikabilitasnya.

Kata kunci: *String Matching, Algoritma Boyer Moore, Pattern*

1. PENDAHULUAN

String matching adalah subjek dalam domain pemrosesan teks yang lebih luas, dengan adanya rangkaian algoritma pencocokan yaitu komponen dasar yang digunakan dalam implementasi perangkat lunak praktis yang ada di sebagian besar sistem operasi [1]. Algoritma boyer moore merupakan algoritma pencarian string pada tahun 1977 yang dianggap sebagai algoritma yang efisien pada aplikasi umum. Tidak seperti pencarian string [2] yang ditemukan sebelumnya [3], mencocokkan karakter dari pattern kanan atau pola yang dicari merupakan karakteristik algoritma ini [4]. Rancangan ini dimulai dengan mencocokkan teks dari kanan, melainkan bukan dari kiri, maka dari itu informasi yang didapat akan lebih banyak [1].

Brute Force Algoritma merupakan pencocokkan pattern dengan pola kata antara 0 dan n-m untuk mendapatkan pattern dalam teks [5]. Dalam mencocokkan karakter terdapat istilah pattern dan teks [6]. Teks adalah objek pola yang akan dicocokkan dengan pattern. Dan pattern objek diinputkan kemudian dicocokkan [7].

Analisa antara Boyer Moore dan Brute force dalam membandingkan mencari word suggestion dengan MPE (Metode Perbandingan Eksponensial) menggunakan data dari kata pada database, dapat disimpulkan Boyer Moore [8] dan Brute Force dapat digunakan dalam pencarian word suggestion [9], untuk penelitian ini dengan membandingkan Boyer Moore [10] dengan Knuth Morris Path di 2 pencarian data diketahui boyer moore lebih cepat dari knuth morris path untuk melakukan pencarian pada database [11].

Namun untuk mendapatkan pencarian kata pada sistem akan dilakukan analisa String Matching pada

pengembangan aplikasi e-submission [12] menggunakan algoritma boyer moore untuk meningkatkan proses pencarian data yang efisien belum banyak di kaji oleh penelitian lain.

Analisa string matching sendiri menggunakan bahasa pemrograman PHP dalam membentuk model string matching untuk menguji parameter-parameter inputan yang mana parameter tersebut akan dicocokkan dengan data source yang ada pada sistem menggunakan model yang telah dibuat.

2. TINJAUAN PUSTAKA

2.1. Analisis

Menganalisis adalah proses melihat atau memeriksa suatu peristiwa menggunakan data untuk memahami situasi sebenarnya. Analisis dilakukan saat melakukan penelitian atau memproses data. Temuan dari analisis dapat membantu dalam pengambilan keputusan dan membantu orang memahami hal-hal dengan lebih baik. Analisis [20] juga dapat membedah suatu topik menjadi bagian-bagian komponennya dan kemudian menganalisis bagian-bagian tersebut untuk mendapatkan informasi yang tepat dan penting.

2.2. String Matching

Sebuah algoritma yang digunakan untuk membandingkan satu teks dengan teks lain disebut pencocokan string (string matching). Boyer-Moore dan metode brute-force adalah dua dari banyak algoritma yang digunakan dalam pencocokan string [21]. Tujuan dari masalah klasik dalam ilmu komputer yang dikenal sebagai "string matching" adalah mencari semua kemunculan pola (sering disebut sebagai "pola") di dalam sebuah string yang lebih panjang (biasanya disebut sebagai "teks"), di mana kedua string

tersebut berasal dari abjad yang sama. Dalam banyak aplikasi, termasuk penambahan data, mesin pencari web, bioinformatika, dan pemrosesan bahasa alami, ini adalah tugas yang penting [22].

2.3. E-Submission

E-Submission adalah file atau dokumen dikirimkan secara elektronik, yaitu melewati e-mail, formulir online di Internet, atau website (Kim et al., 2020). Secara tradisional, manuskrip mengacu pada sesuatu yang khusus "tulisan tangan". Namun, dalam bahasa umum, dan khususnya dalam konteks komputer dan Internet, sebuah manuskrip bahkan dapat mengacu pada dokumen (teks atau lainnya) yang ditulis atau dibuat dengan komputer, dan dapat diperluas untuk mencakup foto dan video digital juga. termasuk survei online. Dengan kata lain, apa pun yang dibuat dan dikirim secara online dapat dianggap sebagai pengiriman elektronik [23].

2.4. Algoritma Boyer Moore

Salah satu metode pencarian string adalah algoritma Boyer Moore, yang dikembangkan pada tahun 1977 oleh Robert S. Boyer dan J. Strother Moore. Metode ini dianggap lebih efektif dalam semua kasus. Algoritma Boyer-Moore mulai mencocokkan karakter-karakter dari sisi kanan string target atau pola. Ini adalah metode yang melakukan pelacakan karakter dalam teks sambil mencocokkan string dengan menggulirkan dari kanan ke kiri. Faktor bahwa algoritma ini mencocokkan dari kanan ke kiri membuatnya lebih mudah dan lebih cepat secara rata-rata dibandingkan dengan teknik lainnya [4]. Sejumlah algoritma yang bertujuan untuk mengoptimalkan peningkatan pergeseran didasarkan pada metode Boyer-Moore. Algoritma Boyer-Moore secara khusus membandingkan karakter-karakter dari pola P dengan karakter-karakter dari teks T dalam arah dari kanan ke kiri. Untuk menghitung nilai pergeseran tertinggi yang mungkin untuk pola tersebut, algoritma ini menggunakan nilai pergeseran dari "Tabel Karakter Baik" (Good Suffix Table) dan "Tabel Karakter Buruk" (Bad Character Table) [10].

3. METODE PENELITIAN

Tahapan dalam melakukan Systematic Literature Review sebagai landasan penelitian, tahapan – tahapan yang terlibat dalam melakukan Systematic Literature Review sebagai berikut:

3.1. Research Question

Table 1. Reseach Question

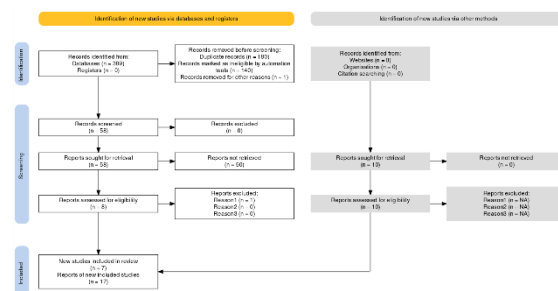
Id	Research Question
Rq1	Apakah ada kaitan pada permasalahan pada penelitian bahwa search engine dapat digunakan dengan metode <i>String Matching</i> ?

3.2. Searching for the Literature

Pengumpulan Artikel atau jurnal dicari dengan menggunakan tools, yaitu Watase Uake untuk membantu dalam pencarian. Watase Uake adalah merupakan sistem online yang dirancang untuk melakukan kolaborasi penelitian antar peneliti. Dirintis sejak tahun 2018 dan mulai dikembangkan dengan melibatkan peneliti dari beberapa perguruan tinggi pada tahun 2020. Watase.web.id dibuat dengan tujuan memfasilitasi peneliti dalam melakukan riset bersama (riset kolaborasi). Dengan menggunakan Watase, berarti peneliti bersedia saling berbagai dengan peneliti lain. Pencarian artikel dengan topik yang terkait dengan menggunakan kata kunci: String Matching, Boyer Moore Algorithm, pattern, Search Engine.

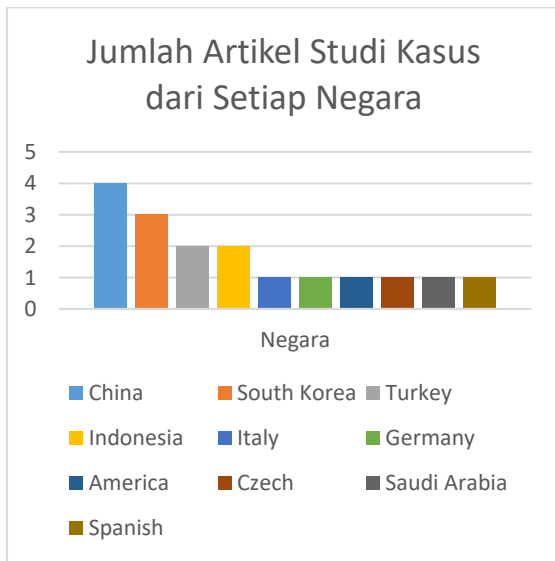
3.3. Selection Criteria

Dari pencarian menggunakan watase Uake mendapatkan 389 artikel. Pemilihan kriteria artikel ada dua, yaitu kriteria inklusi merupakan kriteria yang disesuaikan, serta kriteria eksklusi adalah kriteria yang tidak termasuk atau dikecualikan. Sampel yang diinginkan dari pembahasan ini adalah jangka maksimal dari tahun 2017, lalu berhubungan dengan klasifikasi, dan menggunakan metode. Kriteria eksklusi dari penelitian ini adalah artikel yang diterbitkan sebelum tahun 2017 dan artikel yang tidak memiliki hasil. Dari hasil pencarian terdapat 82 artikel duplikat dan 39 artikel yang tidak termasuk. Selanjutnya terdapat 30 artikel yang tidak sesuai dengan topik serta 30 artikel yang tidak masuk ke periode. Setelah melalui tahapan identification dan screening, artikel yang termasuk adalah sebanyak 18 artikel.

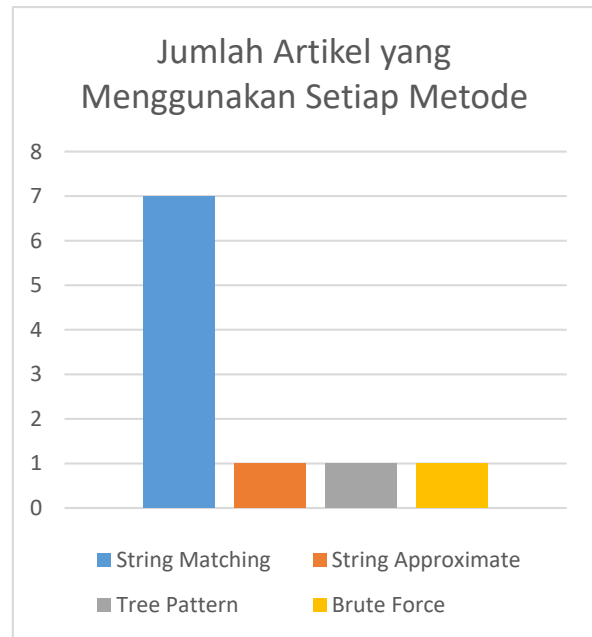


Gambar 1. Selection Criteria

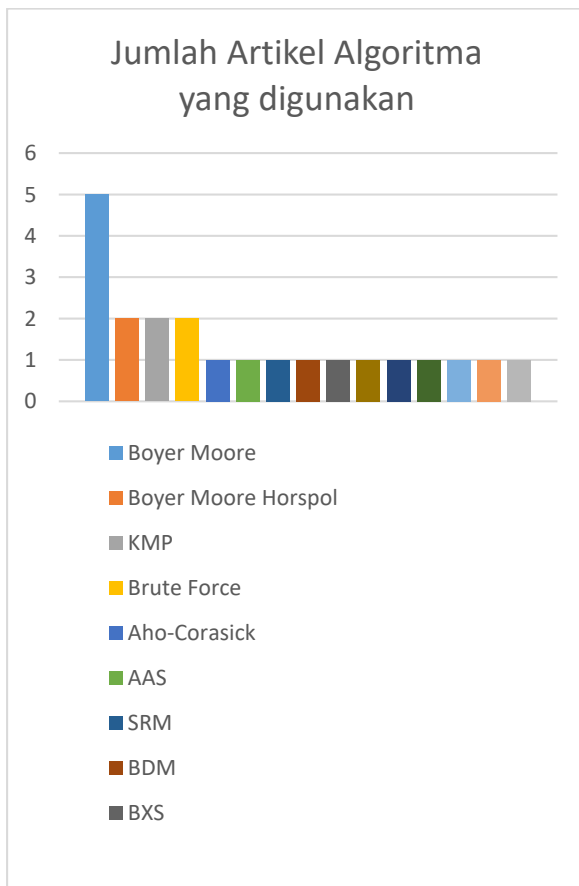
4. HASIL DAN PEMBAHASAN



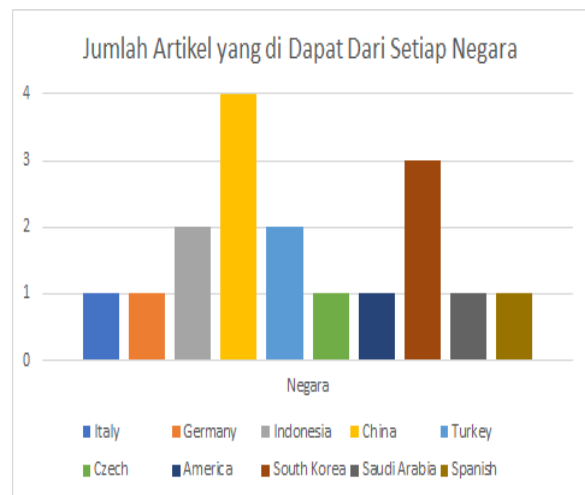
Gambar 2. Jumlah Artikel Studi Kasus dari setiap Negara



Gambar 4. Jumlah Artikel yang menggunakan setiap Metode



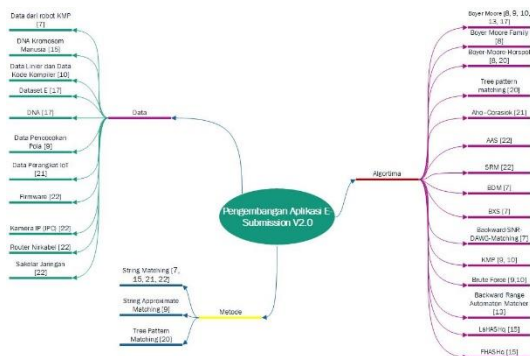
Gambar 3. Jumlah Artikel Algoritma yang digunakan



Gambar 5. Jumlah Artikel yang di dapat dari setiap Negara

Tabel 2. Ringkasan Artikel terpilih

No	Area Penelitian	Karakteristik Data	Metode	Model
1	A weak approach to suffix automata simulation for exact and approximate string matching [14].	Data dari robot KMP	String Matching	Algoritma BDM, Algoritma BXS, algoritma Backward SNRDAWG Matching (BSDM)
2	LsHASHq: A string matching algorithm exploiting longer q-gram shifting [15].	DNA kromosom manusia 1	String Matching	LsHASHq, FHASHq
3	Improving hash-q exact string matching algorithm with perfect hashing for DNA sequences [10].	Dataset E dan DNA	String Matching	Algoritma BoyerMoore
4	Determining the base address of MIPS firmware based on absolute address statistics and string reference matching[3].	12 jenis file firmware, kamera IP (IPC), router nirkabel, dan sakelar jaringan	String Matching	algoritma AAS dan algoritma SRM



Gambar 6. Taksonomi

Adapun pengumpulan data yang digunakan peneliti dengan metode berupa studi dokumen yang mana mengambil data dari sumber data yang telah tersedia. Adapun sumber data yang dimaksud adalah basis data dari aplikasi Esubmission yang menyimpan data tugas pada penelitian sebelumnya. Berikut adalah data yang diperoleh dari data source yang digunakan oleh penulis dalam penelitian:

Tabel 3. Data Source

Tgscode	matkul name	Tgstitle	Tgsdesc
TGS31202301071	algoritma 1	Find duplicate number edit	Cari duplicate value di array berikut edit: {1,2,4,1,2,4,6,4,8,8,9,3,6,7,0}
TGS31202301073	desain interaksi	Create database untuk sistem pembuat ktp edit	Silahkan buat database dengan kriteria – kriteria berikut 1. terdapat minimal 10 tabel 2. harus berelasi 3. memiliki constraint dan indeks edit
TGS31202301164	Pemrograman 1	Laporan Pemrograman	Buat makalah pemrograman

Tgscode	matkul name	Tgstitle	Tgsdesc
TGS3120230116110	Pemrograman 2	Laporan Pemrograman	Buat makalah tentang html
TGS312023122111	Sister	Test Tugas Baru	Buat source code python
TGSD5420230122112	Pemrograman 3	Tugas dengan Datetim e	Buat malakah tentang php

Penjelasan dari Gambar 6 taksonomi literature merupakan kumpulan dari penelitian sebelumnya yang ada dan sudah dikerjakan dan terdapat kaitan pada permasalahan pada penelitian bahwa search engine dapat digunakan dengan menggunakan beberapa metode. Adapun metode yang dapat diusulkan pada penelitian yang akan dijalankan berdasarkan literatur yang diambil yaitu:

Tabel 3. Literature yang digunakan

Metode	Algoritma
String Matching	Boyer Moore

Dari beberapa metode dan Algoritma yang di rangkum dari beberapa referensi tersebut bahwa metode string matching menjadi metode yang bagus dengan dipadukan dengan Algoritma Boyer Moore dengan hasil dari beberapa penelitian terkait bahwa metode dan model tersebut mendapatkan hasil yang dapat dikatakan baik.

Hasil pencarian pencocokan pola dengan menggunakan Algoritma Boyer Moore dapat digunakan untuk mengidentifikasi pola-pola dalam data dalam mendapatkan informasi dengan tujuan meningkatkan pemrosesan pencarian data dengan Algoritma Boyer Moore [4].

Hasil penelitian dan pengujian dari pencarian rangkaian dengan menggunakan metode Algoritma Boyer Moore menunjukkan bahwa program tersebut dinyatakan layak untuk digunakan, karena aturan asosiatif yang dibentuk dan nilai pencarian yang dihasilkan sama dengan output program yang

dikembangkan. Hasil pengolahan pencarian dengan menggunakan algoritma Boyer Moore dapat digunakan untuk mencari pencocokan data untuk menunjang informasi dalam pemrosesan boyer moore dengan pencarian data [4].

hasil dari pembahasan ini yaitu dengan sistem yang dapat menentukan tingkatan pencarian dalam string matching dengan menggunakan algoritma boyer moore untuk menjalankan pencocokan kata dimulai dari karakter kanan kemudian ke kiri berlandaskan nilai dari occurrence heuristic (HO) dan match heuristic (MH) dalam menentukan pergeseran. Hasil dari pengujian data source dengan menginputkan 6 data yang tersedia menunjukkan algoritma boyer moore ini mampu mencocokkan dengan cepat. Meskipun semakin memperolah banyak data semakin membutuhkan waktu yang cukup lama [18].

Berdasarkan penelitian yang dilakukan bahwa dengan string matching dapat membantu proses pengelolaan data menjadi lebih efektif dari sebelumnya, seperti penyimpanan data yang terintegrasi, pengambilan data yang lebih cepat, dan dapat diakses langsung oleh pengguna dan penerapan algoritma Boyer-Moore dapat membantu proses pencarian data menjadi lebih efektif [19].

5. KESIMPULAN DAN SARAN

Kesimpulan Literature Review yang membantu peneliti mengkaji artikel dengan tools watase Uake untuk menemukan metode dan model yang tepat. Sebanyak 389 artikel ditemukan. Dari seluruh artikel yang ditemukan, tersisa 17 artikel yang telah dilakukan identifikasi dan screening yang memenuhi kriteria inklusif dan eksklusif dibuat dengan diagram prisma. Hasil dari tinjauan literatur memperlihatkan bahwa penelitian sebelumnya yang ada dan sudah dikerjakan dan terdapat kaitan pada permasalahan pada penelitian bahwa search engine dapat digunakan dalam beberapa metode yaitu string matching dan Algoritma boyer moore. Selanjutnya pada dataset yang digunakan yaitu dengan dataset peneliti, dikarenakan penelitian ini tentang teknis penggunaan Boyer Moore di String Matching yaitu untuk specific lebih mengarah ke algoritma dan metode, Saran Pada percobaan disediakan benchmark dengan metode lain yang mendukung string matching dan String matching diterapkan pada aplikasi Esubmission v2.

DAFTAR PUSTAKA

- [1] C. and Charras and T. Lecroq, "Handbook of Exact String Matching Algorithms," p. 238, 2004.
- [2] M. E. Hussain, M. K. Gupta, and R. Hussain, "A lightweight string-matching technique for secure communication within IoT energy systems technology," *IoT Enabled Multi-Energy Syst.*, pp. 159–166, Jan. 2023, doi: 10.1016/B978-0-323-95421-1.00007-0.
- [3] X. Zhu, Y. Zhang, L. Jiang, and R. Chang, "Determining the base address of MIPS firmware based on absolute address statistics and string reference matching," *Comput. Secur.*, vol. 88, p. 101504, Jan. 2020, doi: 10.1016/J.COSE.2019.02.015.
- [4] Y. Faqih, Y. Rahmanto, A. A. Aldino, and B. Waluyo, "BULLETIN OF COMPUTER SCIENCE RESEARCH Penerapan String Matching Menggunakan Algoritma Boyer-Moore Pada Pengembangan Sistem Pencarian Buku Online," vol. 2, no. 3, pp. 100–106, 2022, doi: 10.47065/bulletincsr.v2i3.172.
- [5] D. Ramadhani, "Perancangan Aplikasi Pencarian Buku Pada Perpustakaan Islamic International School Darul Ilmi Murni Dengan Algoritma Knuth Morris Pratt," *Pelita Inform. Budi Darma*, vol. XVI, no. 1, pp. 25–27, 2017.
- [6] J. Trávníček, J. Janoušek, B. Melichar, and L. Cleophas, "On modification of Boyer-Moore-horspool's algorithm for tree pattern matching in linearised trees," *Theor. Comput. Sci.*, vol. 830–831, pp. 60–90, Aug. 2020, doi: 10.1016/J.TCS.2020.04.027.
- [7] F. T. Waruwu and P. Ramadhani, "Perancangan Aplikasi Translator Bahasa Indonesia Ke Bahasa Inggris Esp (English for Specific Purpose) Computer Sains Untuk Peningkatan Informasi Mahasiswa Menggunakan Brute Force Method," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 410–414, 2018, doi: 10.30865/komik.v2i1.967.
- [8] Y. Duan, H. Long, and Y. Q. Qu, "Application of Improved BM Algorithm in String Approximate Matching," *Procedia Comput. Sci.*, vol. 166, pp. 576–581, Jan. 2020, doi: 10.1016/J.PROCS.2020.02.017.
- [9] A. Januardi, "Analisa Perbandingan Algoritma Brute Force dan Boyer Moore dalam Pencarian Word Suggestion Menggunakan Metode Perbandingan Eksponensial," *Pelita Inform. Budi Darma*, vol. 4, no. 1, pp. 18–24, 2013, [Online]. Available: http://www.stmik-budidarma.ac.id/Email:andri_januardi@yahoo.comABSTRAK
- [10] A. A. Karcioğlu and H. Bulut, "Improving hash-q exact string matching algorithm with perfect hashing for DNA sequences," *Comput. Biol. Med.*, vol. 131, p. 104292, Apr. 2021, doi: 10.1016/J.COMPBIOMED.2021.104292.
- [11] A. Fau, Mesran, and G. L. Ginting, "Analisa Perbandingan Boyer Moore Dan Knuth Morris Pratt Dalam Pencarian Judul Buku Menerapkan Metode Perbandingan Eksponensial (Studi Kasus : Perpustakaan STMIK Budi Darma)," *J. Times (Technology Informatics Comput. Syst.)*, vol. 6, no. 1, pp. 12–22, 2017.
- [12] Z. Ling, J. Na, S. Yan-Li, and J. Sriyanto, "School culture and professional development of school teachers from urban and rural areas in China," *Cakrawala Pendidik.*, vol. 39, no. 3, pp. 609–619, 2020, doi: 10.21831/cp.v39i3.31580.

- [13] I. Alfarobi et al., "RANCANG BANGUN SISTEM INFORMASI PEMESANAN KOPI DENGAN MENGGUNAKAN METODE WATERFALL Program Studi Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika 2)," *Ijns.org Indones. J. Netw. Secur.*, vol. 9, no. 1, pp. 1–5, 2020.
- [14] S. Faro and S. Scafiti, "A weak approach to suffix automata simulation for exact and approximate string matching," *Theor. Comput. Sci.*, vol. 933, pp. 88–103, Oct. 2022, doi: 10.1016/J.TCS.2022.08.028.
- [15] A. M. Al-Ssulami, A. M. Azmi, H. Mathkour, and H. Aboalsamh, "LsHASHq: A string matching algorithm exploiting longer q-gram shifting," *Inf. Process. Manag.*, vol. 59, no. 5, p. 103057, Sep. 2022, doi: 10.1016/J.IPM.2022.103057.
- [16] C. Ryu, T. Lecroq, and K. Park, "Fast string matching for DNA sequences," *Theor. Comput. Sci.*, vol. 812, pp. 137–148, 2020, doi:10.1016/j.tcs.2019.09.031.
- [17] H. Li and J. Wang, "Collaborative annealing power k-means++ clustering," *Knowledge-Based Syst.*, vol. 255, p. 109593, 2022, doi: 10.1016/j.knosys.2022.109593.
- [18] Ahmad, I., Borman, R. I., Caksana, G. G., & Fakhrurozi, J. (2021). Implementasi String Matching Dengan Algoritma Boyer-Moore Untuk Menentukan Tingkat Kemiripan Pada Pengajuan Judul Skripsi/Ta Mahasiswa (Studi Kasus: Universitas XYZ). *SINTECH (Science and Information Technology) Journal*, 4(1), 53–58.
- [19] 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), 1308-1316.
- [20] N. Navani et al., "The Accuracy of Clinical Staging of Stage I-IIIa Non-Small Cell Lung Cancer: An Analysis Based on Individual Participant Data," *Chest*, vol. 155, no. 3, pp. 502–509, Mar. 2019, doi: 10.1016/J.CHEST.2018.10.020.
- [21] C. Ryu, T. Lecroq, and K. Park, "Fast string matching for DNA sequences," *Theor. Comput. Sci.*, vol. 812, pp. 137–148, Apr. 2020, doi: 10.1016/J.TCS.2019.09.031.
- [22] A. M. Al-Ssulami, A. M. Azmi, H. Mathkour, and H. Aboalsamh, "LsHASHq: A string matching algorithm exploiting longer q-gram shifting," *Inf. Process. Manag.*, vol. 59, no. 5, p. 103057, Sep. 2022, doi: 10.1016/J.IPM.2022.103057.
- [23] Kim, I., Choi, J., Teo, E. A. L., & Sun, H. (2020). Development of kbim e-submission prototypical system for the openbim-based building permit framework. *Journal of Civil Engineering and Management*, 26(8), 744–756. <https://doi.org/10.3846/jcem.2020.13756>.