

Identifikasi Lagu Berdasarkan Lirik Menggunakan Algoritma Boyer-Moore

I Komang Gede Apriana^{a1}, I Gede Arta Wibawa.^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana, Bali
Jln. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, 08261, Bali, Indonesia
¹gedeapriana36@gmail.com
²gede.artha@cs.unud.ac.id

Abstract

Many people want a fast and efficient search method as technology advances. A song search is one example of this kind of search. A song is a collection of sing-along lyrics with rhythms and melodies for many to enjoy. Due to the large number of song lovers, some people are often constrained by the title of the song to be sung. This is caused by one factor, namely only memorizing some of the lyrics of the song to be sung. Given these problems, in this study a solution was developed, namely the application of identifying song titles based on input from the user's lyrics. The algorithm used by researchers in this study is the Boyer-Moore Algorithm, which is considered better in terms of matching substrings in longer texts. The research method used includes literature study, data collection, implementation, and testing. The implementation results show that the system successfully recognizes song titles with high accuracy based on the given piece of lyrics. In conclusion, this study proves that the development of a song title identification system based on snippets of lyrics using the website-based Boyer-Moore algorithm is an effective method. This system can help users recognize song titles based on the snippets of lyrics they remember with high accuracy.

Keyword: song, lyrics, boyer-moore

1. Pendahuluan

Lagu adalah ciptaan seseorang berupa sejumlah syair yang dinyanyikan dengan nada dan makna tertentu. Panjang, nada tinggi, dan nada rendah setiap lagu membuatnya unik. Lagu adalah bentuk seni lain yang menggunakan keindahan musik untuk menyampaikan pikiran dan emosi manusia. Terkadang saat kita ingin menyanyikan sebuah lagu, kita hanya mengingat sebagian liriknya saja dan lupa judul lagunya. Hal ini menjadi tantangan bagi para pecinta musik dalam mencari dan mengidentifikasi judul lagu berdasarkan potongan lirik yang kita ingat. Dalam penelitian ini, peneliti akan mengajukan sebuah solusi dengan memanfaatkan algoritma pencarian substring untuk mengidentifikasi judul lagu berdasarkan potongan lirik yang diinputkan oleh pengguna.

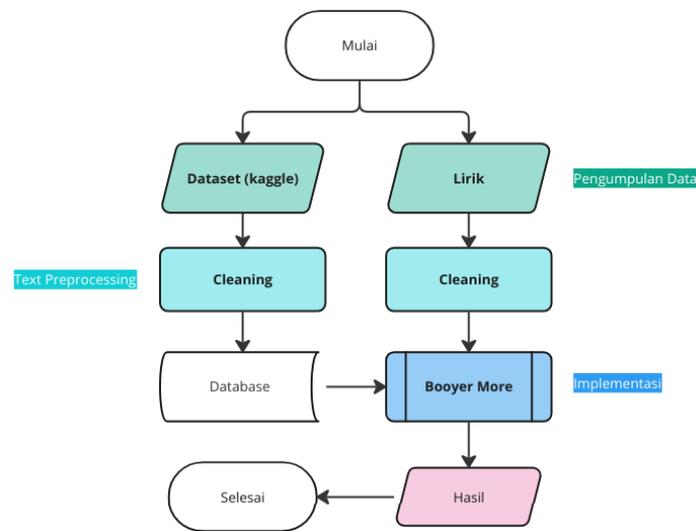
Terdapat berbagai metode atau algoritma yang bisa digunakan untuk pencocokan substring. Diantaranya adalah algoritma *BM (Boyer-Moore)* dan *KMP (Knuth-Morris-Pratt)*. Menurut penelitian sebelumnya, metode Boyer-Moore berperforma lebih baik daripada pendekatan KMP dalam hal pencocokan substring dalam teks yang lebih panjang, terlepas dari kenyataan bahwa setiap algoritma memiliki kelebihan dan kekurangannya sendiri [1]. Metode Boyer-Moore mengambil teknik yang berbeda, memanfaatkan pola yang dicari informasinya untuk melakukan pergeseran yang lebih optimal dan dengan demikian meminimalkan jumlah perbandingan yang diperlukan. Hal ini menjadikan penggunaan metode Boyer-Moore untuk menemukan nama lagu dari sampel lirik yang dimasukkan pengguna dalam penelitian ini sebagai alternatif yang menarik.

Dalam penelitian ini, peneliti mengusulkan untuk mengidentifikasi lagu berdasarkan liriknya menggunakan algoritma Boyer-Moore. Untuk menemukan pola dalam teks, metode Boyer-Moore adalah teknik pencocokan pola yang cepat dan efektif. Metode heuristik algoritma ini

mempercepat proses pencocokan dengan menggunakan data dari pola itu sendiri [2]. Dalam menangani variasi dalam lirik dan perubahan kecil dalam teks lirik, algoritma Boyer-Moore unggul. Peneliti bertujuan untuk mendapatkan hasil identifikasi lagu yang lebih tepat dan efektif berdasarkan dataset lirik lagu yang dikumpulkan dengan menggunakan algoritma Boyer-Moore.

2. Metode Penelitian

Tahapan penelitian yang akan dilakukan oleh peneliti adalah Studi Literatur, Pengumpulan Data, Text Preprocessing, Implementasi, dan Pengujian.



Gambar 1. Alur Metode Penelitian

2.1 Studi Literatur

a. Boyer-Moore

Algoritma Boyer-moore merupakan algoritma yang digunakan pada penelitian ini. Robert S. Boyer dan J. Strother Moore merilis metode Boyer-Moore, salah satu algoritma pencarian string, pada tahun 1977. Algoritma ini dianggap sebagai algoritma aplikasi umum yang paling efektif. Teknik Boyer-Moore mulai mencocokkan karakter dari sisi kanan pola, berbeda dengan algoritma pencarian string yang lainnya [3]. Algoritma ini menawarkan manfaat meminimalkan jumlah perbandingan karakter yang diperlukan selama pencarian pola, terutama ketika pola berbagi sejumlah besar karakter dan selama pencarian pola dalam teks yang panjang. Teknik ini telah banyak digunakan dalam berbagai aplikasi pencocokan dan pencarian string.

Text : **AABAACAADAABAABA**
 Pattern : **AABA**

```

    AABA          AABA
    AABAACAADAABAABA
    0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
                                AABA
    
```

Pattern found at 0, 9 and 12

Gambar 2. Proses Algoritma Boyer-Moore

Algoritma Boyer-Moore memindai pola karakter dari kanan ke kiri dimulai dengan karakter paling kanan, menggunakan mekanisme pencocokan string dari kanan ke kiri. Algoritma Boyer-Moore menggunakan dua fungsi shift: pergeseran akhiran yang baik dan pergeseran karakter buruk untuk mengambil langkah berikutnya setelah ketidakcocokan antara pola karakter dan karakter teks yang sesuai.

b. Python

Bahasa pemrograman tingkat tinggi yang populer dan mudah beradaptasi adalah Python. Python diciptakan pada tahun 1991 oleh Guido van Rossum dengan tujuan menjadi bahasa yang sederhana dan mudah dibaca. Python mendukung paradigma pemrograman fungsional, berorientasi objek, dan terstruktur [4]. Python digunakan secara luas di berbagai industri, termasuk pengembangan aplikasi desktop, pengembangan game, penelitian data, kecerdasan buatan, dan pengembangan web. Python adalah bahasa pemrograman yang sangat membantu untuk analisis data, pembelajaran mesin, dan pemrosesan gambar berkat modul seperti NumPy, Pandas, Matplotlib, dan TensorFlow.

c. Streamlit

Kerangka kerja sumber terbuka yang disebut Streamlit digunakan untuk membuat antarmuka pengguna interaktif (UI) untuk membuat aplikasi data langsung. Framework ini, dibuat oleh Streamlit Inc., memungkinkan pengguna untuk membuat aplikasi web dengan cepat dan mudah menggunakan bahasa pemrograman Python. [5]

2.2 Pengumpulan Data

Dalam penelitian ini, data yang digunakan merupakan informasi lagu yang diambil dari "Spotify Million Song Dataset" yang tersedia di situs Kaggle [6]. Dataset ini terdiri dari ratusan lagu dengan berbagai atribut musik, metadata lagu, dan termasuk lirik. Peneliti mengunduh dataset ini dan menggunakan atribut lirik sebagai bagian dari data yang akan digunakan dalam identifikasi lagu.



#	A Artist	A Title	A Album	# Year	A Date	A Lyric
0	Eminem	Rap God	The Marshall Mathers LP2 (Deluxe)	2013	2013-10-15	Look 1 was gonna go easy on you not to hurt your feelings but I'm only going to get this one chance ...
1	Eminem	Killshot		2018	2018-09-14	you sound like a bitch bitch shut the fuck up when your fans become your haters you done fuckin' bea...
2	Eminem	Godzilla	Music to Be Murdered By	2020	2020-01-17	ugh you're a monster eminem i can swallow a bottle of alcohol and i'll feel like godzilla better h...
3	Eminem	Lose Yourself	The Singles	2002	2002-10-28	look if you had one shot or one opportunity to seize everything you ever wanted in one moment would ...

Gambar 2. Spotify Million Song Dataset

Selain itu, sebagai bagian dari pengujian dan evaluasi algoritma, data uji yang digunakan adalah potongan string lirik yang dimasukkan oleh pengguna. Pengguna akan memberikan potongan lirik lagu yang ingin diidentifikasi berdasarkan algoritma yang diterapkan dalam penelitian ini. Teknik pengumpulan data uji ini memungkinkan pengguna untuk menguji kinerja algoritma dengan menggunakan lirik lagu yang tidak terdapat dalam dataset latih. Dengan demikian, penelitian ini menggunakan sumber data yang akurat dan representatif dalam bentuk dataset lagu yang luas, serta memungkinkan pengguna untuk memberikan data uji berupa potongan string lirik sesuai dengan kebutuhan identifikasi lagu.

2.3 Text Preprocessing (Cleaning)

Data lirik yang akan digunakan dibersihkan setelah data lagu terkumpul. Tanpa mengubah makna dari lirik, tata cara pembersihan ini berupaya menghilangkan tanda baca dari teks lirik. Hanya teks lirik asli yang akan tersisa setelah semua tanda baca, termasuk koma, titik, tanda tanya, tanda seru, dan simbol serupa, telah dihapus. [7] Data pengujian, yang terdiri dari fragmen teks lirik yang dimasukkan pengguna, juga akan melalui prosedur pembersihan serupa. Untuk memastikan bahwa hanya teks lirik murni yang digunakan dalam proses identifikasi lagu, teknik yang sama akan digunakan untuk menghilangkan tanda baca dari teks lirik.

2.4 Implementasi

Teknik Boyer-Moore untuk menentukan judul lagu dari penggalan lirik diimplementasikan menggunakan *Python* sebagai bahasa pemrograman utama. Python dipilih karena kemampuan adaptasinya untuk pemrosesan teks dan ketersediaan berbagai paket untuk pemrosesan data [8]. Untuk meningkatkan interaksi pengguna dengan sistem, peneliti juga menggunakan framework *Streamlit* sebagai antarmuka situs web. Framework ini dipilih karena memungkinkan pembuatan antarmuka pengguna yang interaktif dan mudah dipahami dengan Python. Selain itu, library *Pandas* digunakan untuk membantu dalam manipulasi dan pengolahan data. Dalam implementasi ini, *Pandas* digunakan untuk membaca data lagu dari file CSV, memuat data ke dalam struktur yang sesuai, dan melakukan operasi pemrosesan data seperti filtering dan transformasi [9].

2.5 Pengujian

Pada penelitian ini, jenis pengujian yang dilakukan oleh peneliti adalah pengujian akurasi dari Algoritma Boyer-Moore. Terdapat beberapa metode untuk mengukur akurasi, salah satunya adalah *Confusion Matrix*. *Confusion matrix* adalah tabel yang memperlihatkan jumlah prediksi yang benar dan salah dari setiap kelas.

Tabel 1. Confusion Matrix

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) Type I Error
	0 (Negative)	FN (False Negative) Type II Error	TN (True Negative)

Setelah TP (True Positive), FP (False Positive), FN (False Negative), dan TN (True Negative) didapatkan, maka akurasi dari hasil implementasi akan diukur dengan rumus:

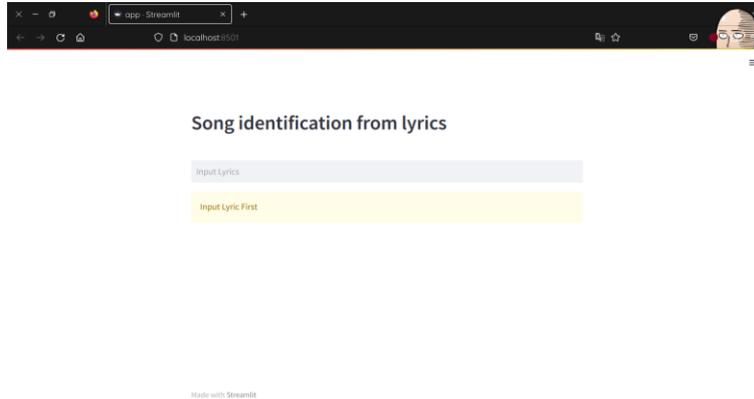
$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

3. Hasil dan Pembahasan

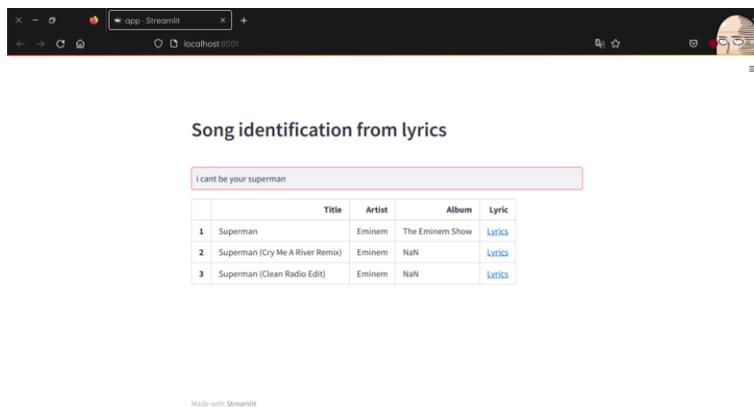
3.1 Hasil Implementasi

Data lagu yang pada tahap proses ini telah diproses dan disimpan dalam struktur data yang sesuai, seperti file CSV (Database), terdiri dari judul lagu, artis, dan lirik. Hal ini dilakukan sebagai bagian dari tahap implementasi. Library *Pandas* kemudian digunakan untuk melakukan proses loading data musik ke dalam aplikasi. Metode ini memudahkan akademisi untuk membaca dan bekerja dengan data musik. Setelah itu, metode Boyer-Moore digunakan untuk membandingkan fragmen lirik yang dimasukkan pengguna dengan lirik lagu database. Teknik pemrograman Python yang sesuai, termasuk penggunaan fungsi string dan array, digunakan untuk membuat pendekatan ini.

Framework Streamlit digunakan untuk membangun antarmuka pengguna yang interaktif dan ramah pengguna. Dalam antarmuka tersebut, pengguna dapat memasukkan potongan lirik lagu yang ingin diidentifikasi judulnya. Selanjutnya, algoritma Boyer-Moore akan dijalankan untuk mencocokkan potongan lirik dengan lirik lagu dalam database, dan hasil identifikasi judul lagu akan ditampilkan kepada pengguna.



Gambar 3. Tampilan Awal Website



Gambar 4. Tampilan Awal Website (Search)

3.2 Hasil Pengujian

Pada tahap pengujian, peneliti menggunakan 50 data potongan lirik dengan label judul dari lagu yang benar, serta menggunakan 600 data lagu yang berasal dari Situs Kaggle yang bisa dianggap sebagai data latih yang telah disimpan di dalam bentuk csv. Jika sistem mengidentifikasi potongan lirik tersebut dengan benar, maka akan dianggap sebagai True Positive. Ini berarti sistem dengan benar mengenali lagu yang sesuai dengan potongan lirik.

Tabel 1. Tabel Hasil Pengujian

Lirik	Judul	Hasil Prediksi	Confusion Matrix
His palms are sweaty, knees weak, arms are heavy	Lose Yourself	Lose Yourself	TP
I'm slim shady, yes I'm the real shady	The Real Slim Shady	The Real Slim Shady	TP
Guess who's back, back again	Without Me	Without Me	TP
I can't breathe, but I still fight while I can fight	Love the Way You Lie	Love the Way You Lie	TP

Lirik	Judul	Hasil Prediksi	Confusion Matrix
I'm not afraid, to take a stand	Not Afraid	Not Afraid	TP
Will the real Slim Shady please stand up?	The Real Slim Shady	The Real Slim Shady	TP
...

Dalam hasil pengujian, ditemukan bahwa dari 50 data potongan lirik yang diuji, semua menghasilkan True Positive (TP). Hal ini menunjukkan bahwa sistem yang dikembangkan berhasil mengidentifikasi judul lagu dengan benar berdasarkan potongan lirik yang diberikan. Keberhasilan ini menunjukkan tingkat ketepatan dan keandalan yang tinggi dari metode yang digunakan, yaitu algoritma Boyer-Moore. Dengan akurasi sebesar 100%, sistem dapat diandalkan dalam mengenali judul lagu dengan tepat berdasarkan lirik yang diberikan oleh pengguna.

4. Kesimpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa pembuatan sistem identifikasi judul lagu berdasarkan potongan lirik dan menggunakan algoritma Boyer-Moore berbasis website dengan bahasa pemrograman Python dan framework Streamlit merupakan metode yang cukup efektif. Berdasarkan lirik yang disediakan, sistem ini dapat mengidentifikasi nama lagu dengan akurasi yang sangat baik. Melalui pengujian yang dilakukan menggunakan dataset lagu yang telah disiapkan, sistem berhasil mencapai tingkat akurasi yang tinggi, menunjukkan keandalannya dalam mengenali judul lagu berdasarkan potongan lirik. Hasil pengujian juga menunjukkan bahwa sistem memiliki performa yang baik dalam menghindari kesalahan pengenalan lagu yang tidak sesuai. Studi ini secara signifikan memajukan bidang identifikasi judul musik dari lirik. Dengan teknik ini, pengguna dapat dengan cepat menemukan nama lagu yang sesuai dengan kata yang masih dapat mereka ingat, sehingga memudahkan mereka mempelajari lebih lanjut tentang musik yang mereka dengar.

Daftar Pustaka

- [1] S. Behera, "Knuth-Morris-Pratt (KMP) vs Boyer Moore Pattern Searching algorithm," [Online]. Available: <https://iq.opengenus.org/kmp-vs-boyer-moore-algorithm/>. Diakses pada Jun. 8, 2023.
- [2] I. Ahmad, R. Indra Borman, G. G. Caksana, and J. Fakhurozi, "Implementasi String Matching dengan Algoritma Boyer-Moore untuk Menentukan Tingkat Kemiripan pada Pengajuan Judul Skripsi/TA Mahasiswa (Studi Kasus: Universitas XYZ)," Sintech Journal, vol. 4, no. 1, pp. 1-10, 2021. [Online]. Available: <https://doi.org/10.3159>
- [3] R. I. Darmawan, A. H. Setianingrum, and Arini, "Implementasi Algoritma Boyer Moore Pada Aplikasi Kamus Istilah Kebidanan Berbasis Web," Jurnal Informatika, vol. 02, no. 01, pp. 1-8, April 2018, ISSN 2579-5341 (online).
- [4] C. R. Severance, "Guido van Rossum: The Modern Era of Python," Computer, vol. 48, pp. 8-10, 2015.
- [5] DataCamp. (<https://www.datacamp.com/tutorial/streamlit>). Diakses pada 11 Juni 2023.
- [6] Kaggle. (<https://www.kaggle.com/datasets/notshrirang/spotify-million-song-dataset>). Diakses pada Mei 23, 2023.
- [7] U. Hasanah et al., "An Experimental Study of Text Preprocessing Techniques for Automatic Short Answer Grading in Indonesian," in 2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE), 2018, pp. 230-234.
- [8] A. G. Karegowda and K. Bharagavi, "Mastering Python Fundamentals with Ease," 2020.
- [9] M. V. Vagizov et al., "Prepare and analyze taxation data using the Python Pandas library," IOP Conference Series: Earth and Environmental Science, vol. 876, p. n. pag., 2021.