Information System and Population Data Search Using the Boyer-Moore Algorithm in Nagari Kotobaru Simalanggang

Fanni Sukma [a,1,*], Ronal Hadi [a], Fitri Chairani [a], Sariani [b], Aditya Wardhani [c]

[a] Department of Information Technology, Politeknik Negeri Padang, West Sumatera, Indonesia

[a] Department of English, Politeknik Negeri Padang, West Sumatera, Indonesia

[a] Department of Electronics Engineering, Politeknik Negeri Padang, West Sumatera, Indonesia

[1] fannisukma@pnp.ac.id
* corresponding author

ABSTRACT

The growth of information technology throughout the world is now faster and stronger than before, one of which is Information Systems. The development of the Nagari information system in Kotobaru Simalanggang aims to increase the efficiency of Nagari administrative staff in searching for population data while making it easier for the public to manage administrative documents such as correspondence. This system enables users to process documents such as letters through a special website. To improve the accuracy of data searches, the Boyer-Moore algorithm was used, with Population Identification Numbers (NIK) and resident names as search criteria. The Laravel framework with Model-View-Controller (MVC) architecture is used to create this website. The waterfall technique is used in the research process. This strategy was chosen because each stage must wait for the preceding stage to be completed before proceeding. In data search, the implementation of the Boyer-Moore method yields a 100% accuracy rate. This guarantees that the public receives the most relevant information.

1. Introduction

Information and communication technology advancements have had a significant influence on many sectors of society, including village government administration. Almost all work operations in the modern day rely on information technology to provide services to the general public. As a result, sustaining information systems becomes critical. The usage of information technology is more frequent in rural regions than in metropolitan ones, particularly in West Sumatra. Local governments should be concerned about this issue so that it can be extended to surrounding cities such as Nagari, which is part of this province [1].

Nagari Kotobaru Simalanggang is a Nagari in Limapuluh Kota District, West Sumatra Province. To improve services to the community in Nagari, a population service information system is needed that focuses on correspondence to manage population data accurately, easily, and quickly. Developing software that is integrated with a population database will be required to create an automated information system for printing letters. This solution will allow users to process letters online via a simple interface. Aside from that, the Nagari guardian personnel will benefit from this information system in controlling the people in Nagari Kotobaru Simalanggang. Implementation of

a data search method using the Boyer-Moore algorithm can ensure an accurate search for population data, thereby increasing efficiency in Nagari administrative management.

A Nagari information system is required to help with communication and demographic data management as a result of these issues. This Nagari information system necessitates the use of a string-matching algorithm, namely one that can match a word. The Boyer-Moore method is employed in texts because the algorithm's principal role is to detect patterns by comparing letters starting from the rightmost of the pattern being searched [2].

The Boyer-Moore algorithm is commonly used in search. Considering the variety of implementations of this technique in various applications to make text searches simpler, numerous text-processing apps, web browsers, and other applications may have used the Boyer-Moore technique in search facilities. Although the actual search can entail text, documents, or files, it can be performed by conducting string matches between keywords and the thing being searched for or against words by referring to a lexical database. The Boyer-Moore algorithm will store the shift information required to do a string search, with the primary difference being that string matching is done from right to left [3]. With this characteristic, when a mismatch occurs in the string comparison, the pattern movement will be directed to go further to avoid character comparisons on strings that are predicted to be incompatible [4].

It is hoped that by implementing string matching algorithms, particularly the Boyer-Moore algorithm, the Nagari information system will have a significant positive impact on improving the quality and speed of services to the community, particularly in managing correspondence and searching for population data. As a result, the installation of this technology is projected to give a long-term contribution to efforts to improve Nagari administration efficiency and enable contact between the community and the government.

2. Literature Review

a. Information Systems

An information system is a computerized system that is managed by an organization, beginning with data collecting and processing and ending with data processing to provide information that is relevant to consumers [5]. Information systems translate raw data into a format that is more valuable to the information's recipient. Data are facts or representations of events like as measurements or observations, whereas information is created by processing the data [6]. Information systems are a phase in the process of converting data into information. An information system is technically described as a set of interconnected components that gather (or receive), analyze, store, and circulate information to support decision-making and monitoring activities inside an organization [7].

b. Laravel

Laravel is an open-source PHP-based web framework that has been extensively accepted by developers all over the world and is used to create online applications. The Model-View-Controller (MVC) architecture principle is used by Laravel [8]. When compared to the generic MVC pattern, the structure of the MVC pattern in Laravel differs slightly. There is a routing mechanism in Laravel that acts as a link between user requests and the controller. Therefore, the controller does not immediately accept the request [9]. The Laravel framework is simple to grasp and simplifies authentication, routing, session management, caching, and a variety of additional Laravel component functions. Laravel also has database migration and unit testing support integration,

making it easier for developers to create complicated applications [10]. The MVC architecture divides the application into components such as data management, controllers, and user interface. The benefits of choosing this approach in development are easier maintenance and scalability.

1. Model, representing the data structure. Usually, the model contains functions that help in database processing.

2. The view is the part that regulates the appearance of the user.

3. The controller is the part that bridges the model and view.

c. Boyer-Moore algorithm

The Boyer-Moore algorithm is a string search algorithm that performs matching from the rightmost side of the string being searched [11]. By using this algorithm, in general, the search process will occur more quickly compared to other algorithms, because matching is done from the right (at the last position of the pattern being searched) [12].

The Boyer-Moore algorithm consists of two core elements, namely the "bad character rule" and the "good suffix rule" [13]. Conditions known as good-suffix shift and bad character shift, are both used to determine the character shift step that will continue in the next comparison [14]. When there are partial matches, the bad character rule is used to move the search to the right (to the front) if there are characters that do not match, while the good suffix rule is used to shift the search to the left (to the rear). Bad character rules use mismatched character information between text and patterns to accelerate searches. To improve searches, good suffix rules use substrings of matching patterns after the text [15]. When compared to other string-matching algorithms, the Boyer-Moore technique is one of the most efficient. Because of the efficient nature of the Boyer-Moore method, various string-matching algorithms based on the principle of the Boyer-Moore algorithm have been created [16].

The stages taken by the Boyer-Moore method when matching strings are executed systematically [17] :

1. The Boyer-Moore algorithm starts matching patterns at the beginning of the text.

2. From right to left, this algorithm will match character by character pattern with characters in the corresponding text, until one of the following conditions is met:

    a. The pattern and text being compared do not match (mismatch).

    b. All characters in the pattern match. The algorithm will notify the discovery at this position.

    c. The algorithm then shifts the pattern by maximizing the good-suffix shift and bad-character shift values, then repeats step 2 until the pattern is at the end of the text.

a) How to calculate Bad-Character Shift (Occurrence Heuristic) and Goos Suffix Shift (Match Heuristic)

To calculate the Occurrence Heuristic and Match Heuristic tables, use the following steps:

Text (S) : FITRI CHAIRANI

Pattern (P): CHAI

Table 1. The application of OH and MH

Pattern (p)	C	H	A	I
Occurrence (OH)	3	2	1	0
Match Heuristic (MH)	4	4	4	1

a. Bad Character Shift (Occurrence Heuristic)

   1. Perform an enumeration from the final position of the string to the first position, beginning with the number 0 since it is the last distance, and record the character encountered (in this case, the character "I").

   2. Retreat to the previous position, and the counter value is increased by 1, if the character in this position has never been found, then the shift value is the same as the counter value (in this example, the character "A" has never been found so the shift value is the same as the counter value. So on until the starting position of the string.

b. Good Suffix Shift (Math Heuristic)

   1. If the character in position 4 is not "I", then shift 1 position, applies to all patterns searched.

   2. If the character "I" matches, but the character before "I" is not "A", then shift it 4 positions, so that the pattern position is past the text.

   3. If the character "AI" matches, but the character before "AI" is not "H", then shift it 4 positions, so that the pattern position is past the text.

   4. If the character "HAI" matches, but the character before "HAI" is not "C", then shift 4 positions.

   5. Next, the most likely and safest shift in the Match Heuristic table is a shift of 4 positions.

b) Stages in the Boyer-Moore Algorithm Process

Manual calculations on the Boyer-Moore Algorithm are used to create a rough idea of how the algorithm works.

Text: FITRI CHAIRANI

Pattern: CHAI

Table 2. Experiment 1 Boyer-Moore Algorithm

Text	F	I	T	R	I		C	H	A	I	R	A	N	I
Pattern	C	H	A	I										

Step I: In the first shift, the character "I" in the pattern does not match the character "R" in the text, so the next shift is based on the value from the OH table. In the OH table, the character "R" is not in the table, so the next shift is as many as the number of characters "I" in the pattern, namely 4.

Table 3 Experiment 2 Boyer-Moore Algorithm

Text	F	I	T	R	I		C	H	A	I	R	A	N	I
Pattern							C	H	A	I				

Step II: In the second shift, the character "I" in the pattern does not match the character "H" in the text, so the next shift is based on the value from the OH table. In the OH table the character "H", is in the table, so the next shift is as many as the number of "H" characters in the OH table, namely 2.

Table 4 Experiment 3 Boyer-Moore Algorithm

Text	F	I	T	R	I	C	H	A	I	R	A	N	I
Pattern						C	H	A	I				

Step III: In the third shift of characters, the "I" in the pattern matches the character "I" in the text. Thus, the search shift is successful because the pattern is found and matches the string.

3. Methodology

The methodology used was the Waterfall method. The Waterfall method, often known as the waterfall model, is a traditional, methodical life cycle model used in the development of software. The waterfall technique is a methodology for the methodical and sequential development of information systems [18]. Winston Royce established the waterfall model in 1970, hence it is frequently regarded as an old approach; nonetheless, it remains the most extensively used model in Software Engineering (SE) [19]. It is called a "waterfall" because each step must wait for the previous stage to be completed before proceeding to the next. This development approach has a linear nature from the beginning of the system development stage, namely the planning stage, to the end of the system development stage, namely the maintenance stage [20]. Subsequent steps cannot be started before previous steps have been completed and it is not possible to go back or repeat previous steps.
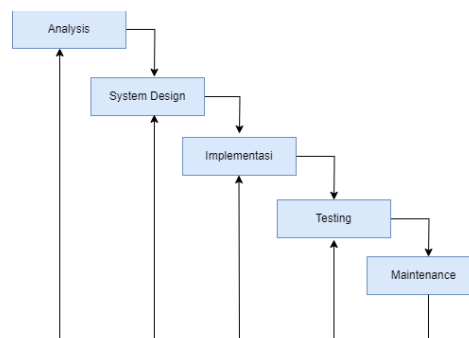

Figure 1. Waterfall Method

a.   Requirements Analysis

The analysis of system requirements is performed at this stage by identifying the required specifications. An approach is also used to comprehend the overall system needs. This analysis seeks to identify existing issues and outline the requirements for developing a website-based Nagari Information System. In addition, during the analysis stage, literature studies are conducted by referring to books, journals, and other relevant sources of information. This literature review covers the fundamental ideas of Laravel, such as defining user access privileges, data management processes, and data search, which will be used in the final project.

b.   System Design

This stage aims to provide a comprehensive overview of the tasks that must be completed and how the desired system will look. This process includes preparing a UML (Unified Model Language)

model by applying Draw IO as a tool to support system planning in the form of UML diagrams. It also includes designing system functionality as well as database design. To describe the system, Balsamiq Mockups are used as a tool to help design the interface in the system.

c. Implementation

This stage begins with preparing all the requirements needed to develop the Nagari Information System in Kotobaru Simalanggang by applying the Boyer-Moore Algorithm to search for population data using calculation algorithms. This step includes creating a database with MySQL, as well as system construction using the Laravel framework as a container for the PHP programming language. For the coding process, the Visual Studio Code application was used as the tool chosen for writing code (code editor).

d. Testing

At this stage, testing is performed to ensure that the system was built by the design that was produced before building the system and to identify the system's vulnerabilities. This testing will help to avoid flaws and errors in the system. Black box testing is used, which focuses on verifying the functionality or behavior of the product without regard for the internal implementation or source code structure. The Boye-Moore algorithm will be used for system testing and data search testing at this level. This testing stage is only reached via the method utilized.

4. Results

4.1. Nagari Information System and Application of the Boyer-Moore Algorithm

The Nagari Kotobaru Simalanggang Information System is an endeavor to upgrade the administrative system so that the community can receive better and more efficient services. The Nagari Kotobaru Simalanggang Information System intends to provide the public with a realistic alternative to printing letters. The Boyer-Moore method is used in this information system to search for population data. The Population Identification Number (NIK) and the name of the resident are used in the data search.

a. Login Page

This display is a display of the login page of the system. This page is used by the admin by entering the email and password that has been registered with the system to access the system.
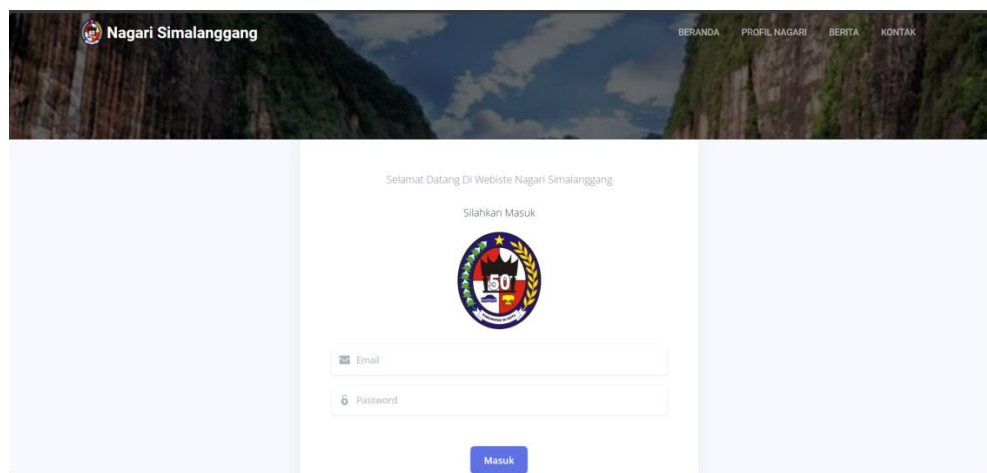

Figure 2. Admin Login Page

b.   Manage Residents page

Data where admins can add data, edit data, and delete data. On this page, the Boyer-Moore algorithm is applied in searching for population data. Searching for population data can use the resident's ID or name.
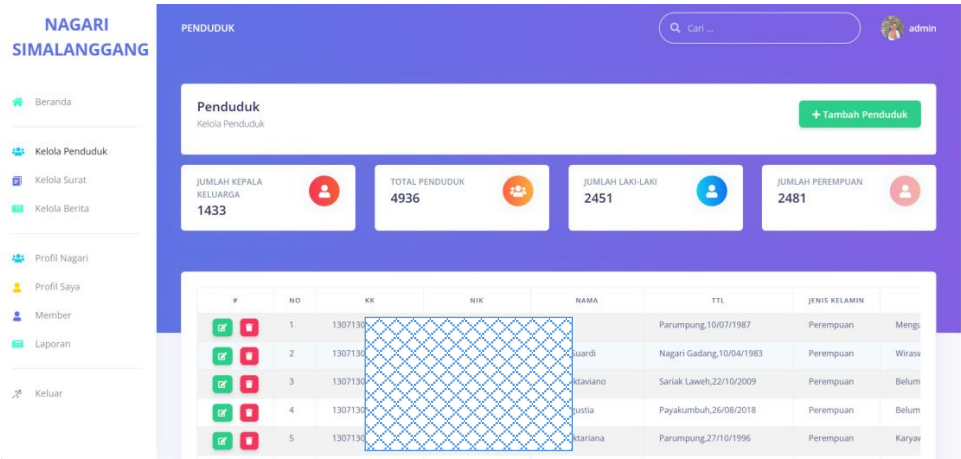

Figure 3. Manage Population Pages

c.   Application of Boyer's Algorithm

Data searches can be performed using a nickname or name, as shown in the image below. The speed of searching for data, whether using a nickname or name, is listed at the bottom of the table.
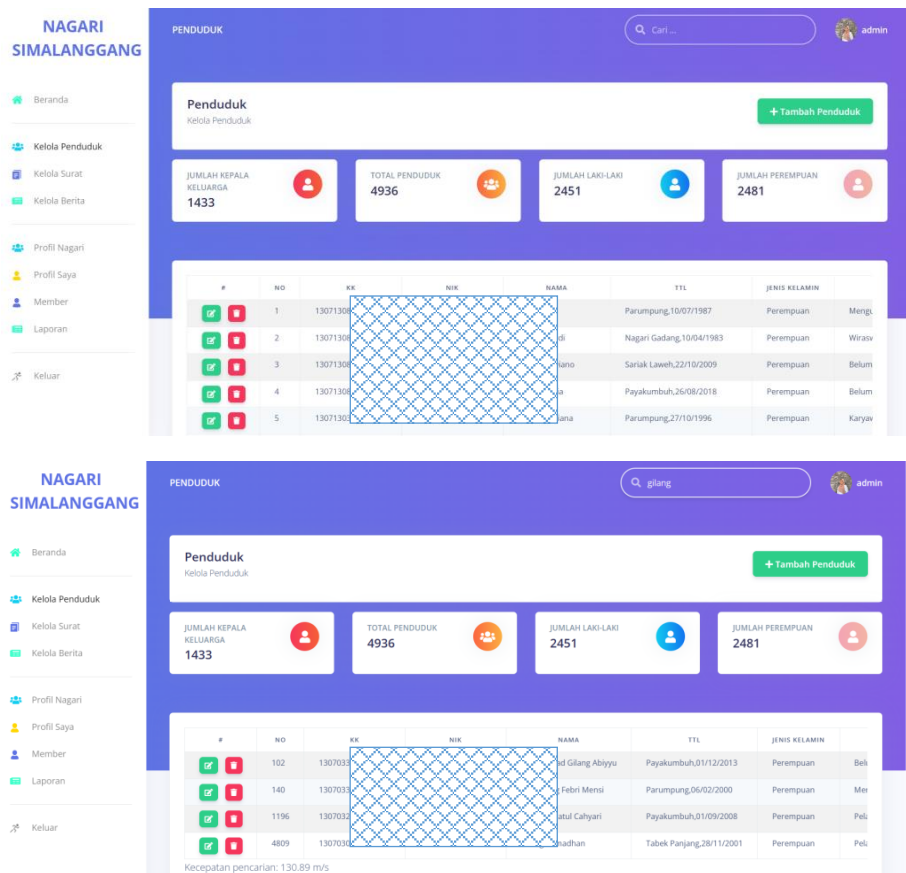

Figure 4. Application of the Boyer-Moore Algorithm

Fanni Sukma et.al (Information System and Population Data Search Using the Boyer-Moore Algorithm …)

d.   Login User

This is a display of the system's login page. Users access the system by entering the ID, email, and password that they have registered with the system on this page.
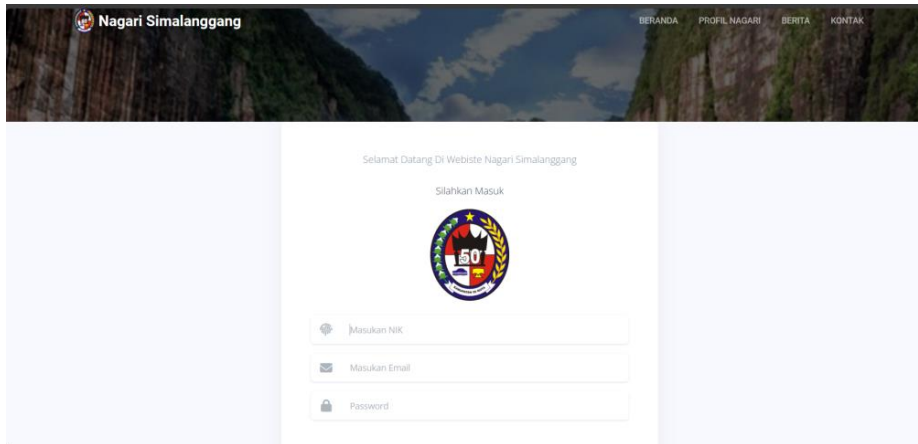

Figure 5. User Login Page

If a person does not have an account, they will be directed to this page. Users must enter their name, nickname, email address, and password. You cannot register if the NIK entered is not in the demographic statistics, or if you are not from Nagari Kotobaru Simalanggang.

Following registration, the user must wait for confirmation from the operator. If the registration is successful, the user will be able to log in.
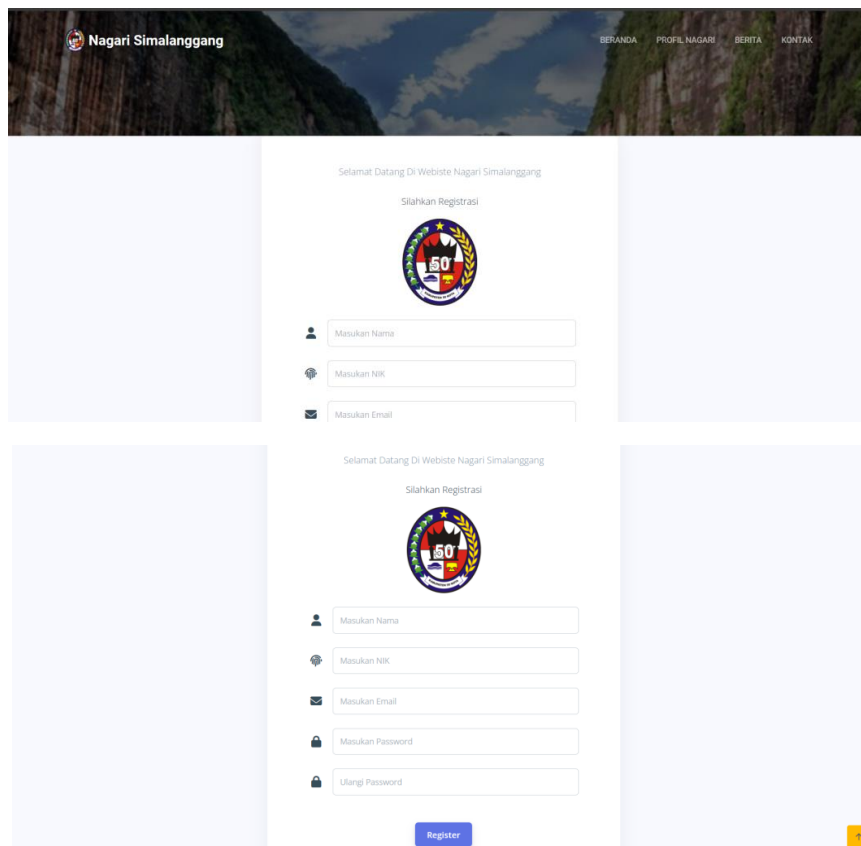

Figure 6. User Register Page

e. Letter

The letter page is a page that the user will access which contains letters that have been input by the admin so that the user can see the letters in this information system. Users can select letters to print. This page will appear if the user has logged in. The user has to log in before they may control mail.
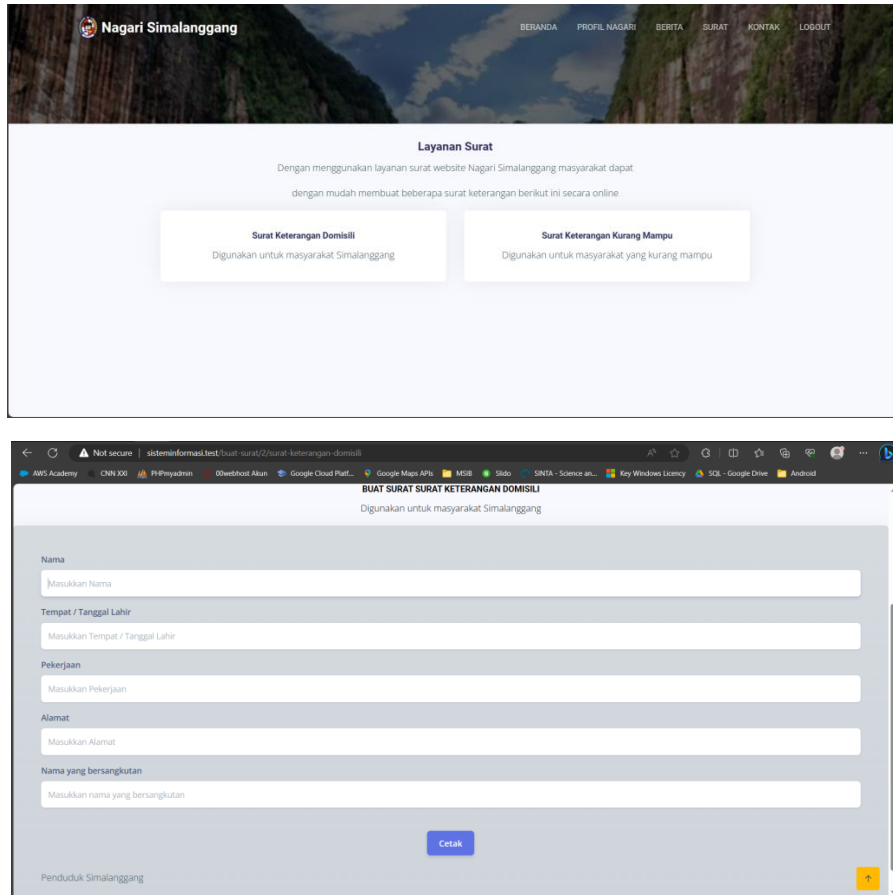

Figure 7. Letter Page

4.2. Results Analysis and Testing

In the Nagari Kotobaru Simalanggang Information System, the Boyer-Moore method is used. This approach is used to find data in the system. The Boyer-Moore Algorithm was tested during the data search process to determine the system's correctness in producing the desired data. Input the resident's ID or name to do a data search. The amount of data used was 4933 Nagari Kotobaru Simalanggang inhabitants.

The accuracy formula in the Receiver Operating Characteristic is a visual tool for measuring the performance of a model. Only the ID field and the name of the resident are used to search the data.

$$\textbf{Accuracy} = \frac{\sum \textit{True Positive } (TP)) + \sum \textit{True Negative } (TN)}{\sum \textit{Total Population}} \tag{1}$$

Formula Description:

True Positive (TP) = Correct search results

True Negative (TN) = Data that is not selected correctly

Total Population = Number of data tested

The trial was carried out 5 times using the NIK and the name of the resident. Accuracy calculations use the accuracy formula in the Receiver Operating Characteristic, to search for data only based on the NIK and name fields.

Table 5. Boyer-Moore Algorithm Testing

No	Searched Words	The Correct Data Found (TP)	The Wrong Data Found	The Unselected Data (TN)	Time	Accuracy
1	Nella Gusri Purnama Sari	1	0	4932	63.14 m/s	Accurate 100%
2	13071322100 90003	1	0	4932	57.03 m/s	Accurate 100%
3	Putri	41	0	4892	51.98 m/s	Accurate 100%
4	Limda	1	0	4932	49.37 m/s	Accurate 100%
5	9091	8	0	4925	51.71 m/s	Accurate 100%
The Average Search Speed					54.64 m/s / 0.05464 s	

Searches in the test table above may be done using NIK and the name of the resident. In the first experiment, a four-word name was utilized with a search time of 63.14 m/s. In the second experiment, NIK (16 digits) was utilized with a search time of 57.03 m/s. The third experiment, named Putri, had a search time of 51.98 m/s. The fourth experiment employed a part of Limda's name, Halimda, with a search time of 49.37 m/s. In the fifth experiment, a piece of NIK (4 digits) with a search speed of 51.71 m/s was employed.

The test table above produces an average search time of 54.64 milliseconds or 0.05464 seconds with 4933 data. From 5 trials the total accuracy obtained is 100%.

5.　　Conclusion

Following are the conclusions that can be derived from the findings of the research:

1. The Nagari Kotobaru Simalanggang Information System is an endeavor to modernize the administrative system so that the community may receive better and more efficient services.

2. The Nagari Kotobaru Simalanggang Information System intends to offer the public a realistic alternative to printing letters.

3. The application of the Boyer-Moore algorithm to search population data using NIK and resident names.

4. Create an application that is coupled with a database to guarantee that all system implementation information is effectively saved in the database.

5. The findings of the study utilizing the Boyer Moore Algorithm generated 100% accuracy with an average search time of 54.64 milliseconds or 0.05464 seconds after 5 tries to search population data.

Acknowledgment

References

[1]   F. Sukma, E. Asri, Y. Yuhefizar, and U. Asmi, "Pelatihan Tata Kelola Sistem Informasi Untuk Nagari Binaan Politeknik Negeri Padang Di Nagari Ranah Pantai Cermin," *J. Abdimas Pengabdi. dan Pengemb. Masy.*, vol. 3, no. 1, pp. 1–5, 2021.

[2]   E. H. E. Sihombing, I. K. Jaya, and S. Sitepu, "Penerapan Model View Controller (MVC) Pada Perancangan Aplikasi Pencarian Pada Sistem Informasi Perpustakaan Berbasis Android Menggunakan Algoritma Boyer Moore (Studi Kasus: Perpustakaan SMA Methodist 1 Medan)," *Methotika J. Ilm. Tek. Inform.*, vol. 1, no. 2, pp. 52–59, 2021.

[3]   B. Harpad and S. Salmon, "Penerapan Algoritma Boyer-Moore Dalam Perpustakaan Virtual SMK Muhammadiyah 1 Samarinda Sebagai Pendukung Pembelajaran Daring," *J. Ilm. Matrik*, vol. 23, no. 2, pp. 182–188, 2021.

[4]   Y. Faqih, Y. Rahmanto, A. A. Aldino, and B. Waluyo, "Penerapan String Matching Menggunakan Algoritma Boyer-Moore Pada Pengembangan Sistem Pencarian Buku Online," *Bull. Comput. Sci. Res.*, vol. 2, no. 3, pp. 100–106, 2022.

[5]   R. Setiyanto, N. Nurmaesah, and N. S. A. Rahayu, "Perancangan Sistem Informasi Persediaan Barang Studi Kasus di Vahncollections," *J. Sisfotek Glob.*, vol. 9, no. 1, 2019.

[6]   E. Wulandari, "Sistem Informasi Peminjaman Dan Pengembalian Buku Pada Perpustakaan Nagari Saning Bakar Berbasis Web," *Indones. J. Technol. Informatics Sci.*, vol. 1, no. 2, pp. 37–42, 2020, doi: 10.24176/ijtis.v1i2.4889.

[7]   S. Hasan and N. Muhammad, "Sistem Informasi Pembayaran Biaya Studi Berbasis Web Pada Politeknik Sains Dan Teknologi Wiratama Maluku Utara," *IJIS - Indones. J. Inf. Syst.*, vol. 5, no. 1, p. 44, 2020, doi: 10.36549/ijis.v5i1.66.

[8]   W. N. WK and E. Pangestu, "Perancangan Sistem Informasi Akuntansi Persediaan Bahan Baku Daging Berbasis Web Menggunakan Laravel Dan Heidisql Pada Pt. Kirana Semesta Pangan," *J. Tekknologi Inf.*, vol. 2, no. 2, pp. 1–10, 2022, [Online]. Available: http://jurnal.lpkia.ac.id/index.php/jti/article/view/408

[9]   D. Purnama Sari and R. Wijanarko, "Implementasi Framework Laravel pada Sistem Informasi Penyewaan Kamera (Studi Kasus di Rumah Kamera Semarang)," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 1, p. 32, 2020, doi: 10.36499/jinrpl.v2i1.3190.

[10]   R. Somya and T. M. E. Nathanael, "Pengembangan Sistem Informasi Pelatihan Berbasis Web Menggunakan Teknologi Web Service Dan Framework Laravel," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 51–58, 2019, doi: 10.33480/techno.v16i1.164.

[11]   B. Gozali, Y. C. Giap, and D. Lasut, "Perancangan Aplikasi Kamus Istilah Komputer Dengan Metode Algoritma Boyer Moore Berbasis Android," *Algor*, vol. 3, no. 1, pp. 38–49, 2021, doi: 10.31253/algor.v3i1.660.

[12]   M. Aulia, "Penerapan Algoritma Boyer Moore Untuk Pencarian Data Member Pada PT. Boenk Cosmetic Manufacture Berbasis Desktop," *J. Sist. Komput. dan Inform.*, vol. 1, no. 3, pp. 235–238, 2020.

[13]   D. Ariyadi, I. P. Astuti, A. F. Cobantoro, and K. N. Fitri, "Pengarsipan Dokumen Di Kua Punung Kab . Pacitan Jawa Timur Boyer Moore Algorithm Implementation On Document Archiving In Kua Punung Kab . Pacitan Jawa," vol. 6, no. 2, pp. 113–120, 2021.

[14]   N. Buslim, D. Faradita, A. S, and S. Masruroh, "Mobile Based Application 'Hadith for Women'

Using Algorithm Boyer Moore," 2020, doi: 10.4108/eai.7-11-2019.2294564.

[15]　M. Fazira, "Perbandingan Algoritma Knuth-Morris-Pratt Dan Boyer Moore Dengan Metode Perbandingan Eksponensial Pada Aplikasi Kamus Bahasa Indonesia–Jerman Berbasis Android," *Inf. dan Teknol. Ilm.*, vol. 6, no. 2, pp. 202–206, 2019.

[16]　F. Fitriyah, W. R. Putra Gayo, A. Nur Handayani, A. P. Wibawa, and F. Kurniawan, "The implementation of the boyer-moore algorithm in web-based computer and informatic terms dictionary," *4th Int. Conf. Vocat. Educ. Training, ICOVET 2020*, pp. 213–217, 2020, doi: 10.1109/ICOVET50258.2020.9230344.

[17]　M. A. Setiawan, S. Andryana, and A. Gunaryati, "Penerapan Algoritma Boyer Moore Dalam Pencarian Barang Hilang pada Aplikasi FindIt Berbasis Android," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 945, 2021, doi: 10.30865/mib.v5i3.3093.

[18]　G. W. Sasmito, "Penerapan metode Waterfall pada desain sistem informasi geografis industri kabupaten Tegal," *J. Inform. J. Pengemb. IT*, vol. 2, no. 1, pp. 6–12, 2017.

[19]　A. A. Wahid, "Analisis metode waterfall untuk pengembangan sistem informasi," *J. Ilmu-ilmu Inform. dan Manaj. STMIK, no. Novemb.*, pp. 1–5, 2020.

[20]　S. Supiyandi, M. Zen, C. Rizal, and M. Eka, "Perancangan Sistem Informasi Desa Tomuan Holbung Menggunakan Metode Waterfall," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 274, 2022, doi: 10.30865/jurikom.v9i2.3986.