

# Performance Comparative Analysis Boyer Moore, Horspool, Zhu Takaoka Algorithms on the Indonesian Translation Hadith Bukhori

1<sup>st</sup> Nindy Raisa Hanum<sup>1</sup>, 2<sup>nd</sup> Imam Marzuki Shofi<sup>1</sup>, 3<sup>rd</sup> Siti Ummi Masruroh<sup>1</sup>, 4<sup>th</sup> Muhammad Taufik Rusydi<sup>2</sup>, 4<sup>th</sup> Muhammad Afied Hambali<sup>2</sup>  
{nindy.raisa14@mhs.uinjkt.ac.id<sup>1</sup>, imam@uinjkt.ac.id<sup>1</sup>, ummi.masruroh@uinjkt.ac.id<sup>1</sup>,  
taufik@unsa.ac.id<sup>2</sup>, muhafied@unsa.ac.id<sup>2</sup>}

UIN Syarif Hidayatullah, Department of Informatics, Jakarta<sup>1</sup>, University of Surakarta, Faculty of Law, Surakarta, Indonesia<sup>2</sup>

**Abstract.** Along with the times, technological advances are increasing, especially information technology. This encourages the emergence of new innovations in the presentation of digital information. String matching (string matching) is a necessity in presenting information, especially in text search. Therefore, a string matching algorithm is needed which is used to search for one or more strings in a set of strings (text). This research was conducted to compare the performance of several String Matching Algorithms, such as Boyer Moore, Horspool, and Zhu Takaoka Algorithms in the string search process. To assess the performance of the algorithm, this study uses runtime parameters and memory consumption. Both parameters must have the smallest possible value to get the best performance value. The input parameters in this study are several characters and a number of words divided into several scenarios and outputs from this research are runtime and memory consumption. Both parameters are calculated using the Exponential Comparison Method (MDE) to find out the best results. The final result of this study shows that the Boyer Moore algorithm shows the best memory consumption and the Zhu Takaoka Algorithm produces the best runtime. And the calculation results Rank Exponential Comparison Method The algorithm shows Boyer Moore in the first position, then Horspool, and Zhu Takaoka in the last position.

**Keywords:** Algorithm Analysis, Boyer Moore Algorithm, Horspool Algorithm, Zhu Takaoka Algorithm, Runtime, Memory Consumption, Simulation Method.

## 1 Introduction

Every religion has its own guidelines in giving instructions to its adherents. As with the religion of Islam, the Qur'an is a guide to living life in the world and in the hereafter. In addition to the Al-Qur'an, Islam also has a Hadith as a source of law that has a second position in the source of Islamic religion after the Qur'an. Therefore, the Hadith must be believed by Muslims, because the hadith is the words and deeds of the Prophet Muhammad that can be used as a foundation.

In addition to creating innovation in the presentation of information, string matching (string matching) becomes a necessity in information processing, especially in text search. So that by using one or more keywords from the contents of the hadith, can quickly obtain information about the hadith. Therefore, a string matching algorithm is needed which is used

to search for one or more strings in a set of strings (text). Algorithm is a systematic arrangement of steps to resolve a problem [1].

According to Kurniadi, A good algorithm is a step that produces an effective output in a relatively short time and memory usage is relatively little (efficient) with finite steps and the procedure ends either in a situation obtained by a solution or there is no solution. Algorithms for string search have been growing day by day. String search algorithms are considered to have the best results in practice, namely algorithms that move to match strings from right to left. The Boyer-Moore algorithm is one example of an algorithm that uses directions from right to left [2].

In a journal entitled Implementation of the Zhu-Takaoka Algorithm on the Android-Based Music Terms Dictionary Application explains that the Zhu-Takaoka Algorithm is a development of the Boyer-Moore Algorithm. [3] In addition, in the journal entitled Comparison of the Use of the Boyer-Moore Algorithm and Horspool Algorithm on String Search in Media Language explained that the Horspool Algorithm was a distinction from the Boyer-Moore Algorithm, because the results of a researcher named R. Nigel Horspool suggested the added idea of the Algorithm Boyer-Moore [4].

So the author makes a study that discusses the performance of the Boyer Moore Algorithm and the development of the algorithm, namely the Hoorspol Algorithm and the Zhu Takaoka Algorithm on the Indonesian translation of the hadith repository with the case study of Bukhori Hadith. After that, the author will compare the performance of the three algorithms when doing the word search process based on the processing time speed in one process (runtime) and the amount of memory (memory consumption) used. In addition, the author will present the exactness of the word sought in the processing of each algorithm. Therefore, the author gives the title of this study "Comparative Analysis of Boyer Moore, Horspool, and Zhu Takaoka Algorithms on Indonesian Translation Hadith Bukhori Repository".

## 2 Related Works

There are four studies used in this study. First, research conducted by Ditya Gerald, Eko Sedyono, and Yos Richard (2016) conducted a search for three String Matching Algorithms, namely Brute Force Algorithm, Knuth Morris Pratt Algorithm, and Boyer Moore Algorithm to find text errors based on Android with time as parameters [5]. Furthermore, the second study, namely research conducted by Alwin Fau, Mesran, and Guidio Leonarde (2017), they conducted a study comparing Boyer Moore Algorithm and Knuth Morris Pratt Algorithm (KMP) using the exponential select (MPE) method. The calculation that becomes the criterion of the MPE is the amount of memory used and the amount of time needed from each process [6]. The third study, namely the research conducted by Adhi Kusnadi and Abraham Khrisnandi Wicaksono (2017), they analyzed the Horsewater Algorithm accurately and Zhu-Takaoka on searching desktop-based strings by measuring the performance of the two algorithms. The performance used is time in searching patterns in text files [7]. The last is research, conducted by Handrizal, Andri Budiman, Desy Rahayu (2017), they do an accurate analysis of the Zhu Takaoka Algorithm and the Knuth Morris Pratt Algorithm that is implemented in the dictionary based on android with time as a parameter processing. By performing performance measurements, it can be used measurements that can be used, and the accuracy of the words searched in one process from each algorithm. Android with times as parameter processing [8]. The last research, we use Repository developed by Atqia Aulia, et.al (2017), WatsaQ : Repository of Al hadith in Bahasa (Case Study: Hadith Bukhori)[9] [11].

### 3. Simulation Method

In this research, we use simulation method which consist of: problem formulation, conceptual model, input/output data, modelling, simulation, verification and validation, experimentation. [10]

#### 3.1 Problem Formulation

After collecting data through literature study, the writer can formulate a problem, namely performance comparison analysis on the use of Indonesian translation of hadith repositories using Boyer Moore Algorithm, Horspool Algorithm, and Zhu Takaoka Algorithm. This aims to provide a better solution in the matching string process in the repository.

#### 3.2 Conceptual Model

At this stage, the author makes the concept model which is classified into two, namely the first is the internal concept of the Boyer Moore, Horspool, and Zhu Takaoaka algorithms, the second is the concept of the simulation process. The first concept describes an example of a search that occurs in the internal process of the Boyer Moore Algorithm and the Boyer Moore-Horspool Algorithm manually in a word. For the second concept, it explains how the first concept is packaged based on the formulation of the problem generated from the previous stage into a simulation that will be carried out later. In this case, the author describes the flow of the word search process that will be run to get the calculation of the processing time involved in the process of matching strings and data retrieval processing time.

#### 3.3 Input / Output data

The next step is to determine which input will be processed and the output to be obtained. The input needed in this simulation is the hadith keyword that you want to look for in the repository. The output obtained in this simulation is the contents of the hadith, runtime data memory consumption, and the accuracy of the searched word.

#### 3.4 Modeling

At this stage, the author determines the scenario model to be used in the simulation phase. There are four scenario models based on a combination of four data values inputting the size of the pattern. Each algorithm is good for the Boyer Moore Algorithm, Horspool Algorithm, and the Zhu Takaoka Algorithm, each of which carries out four scenarios.

Table 1. Modelling

Scenario	Range of Characters	Output	Examples
1	1-10 Characters	Runtime, Memory Consumption, Accuracy	Wahyu
2	10-20 Characters	Runtime, Memory Consumption, Accuracy	Umar Bin Khatab
3	20-30 Characters	Runtime, Memory Consumption, Accuracy	Tanda-Tanda Orang Munafik

4	30-40 Characters	<i>Runtime, Memory Consumption, Accuracy</i>	Semua Perbuatan Tergantung pada Niatnya
5	40-50 Characters	<i>Runtime, Memory Consumption, Accuracy</i>	Diantara tanda-tanda kiamat adalah diangkatnya ilmu
6	50-60 Characters	<i>Runtime, Memory Consumption, Accuracy</i>	Pertanyaan Jibril kepada Nabi SAW tentang iman, islam, ihsan
7	60-70 Characters	<i>Runtime, Memory Consumption, Accuracy</i>	Bagaimana permulaan wahyu yang diturunkan kepada Rasulullah?
8	70-80 Characters	<i>Runtime, Memory Consumption, Accuracy</i>	Seorang muslim adalah orang yang membuat muslim lain selamat dari lidah dan tangannya

### 3.5 Simulation

At this stage, the system will be run to simulate the performance of each algorithm according to the concept and also the predetermined scenario. And the simulation results will be recorded and then the verification phase will be carried out.

### 3.6 Verification and Validation

At this stage, the author verifies and validates the simulation that has been carried out in the previous stage. Verification is done to ensure that there is an error (error) that occurs when calculating the processing time in Boyer Moore, Horspool, or Zhu Takaoka Algorithms while validation is done to ensure the suitability of the simulation based on the concept model with the formulation of the problem that has been made. If the validation is not fulfilled, then the writer returns to the stage of the concept model to create a new fronting model.

### 3.7 Experimentation

At this stage, the author conducts experiments according to the scenario model created during the modeling stage. Each scenario model is carried out each experiment four times in each algorithm.

- Scenario 1

**Table 2.** Scenario 1

Output	Boyer Moore	Horspool	Zhu Takoka
RunTime (s)	0,158	0,084	0,138
Memory Consumption (kb)	110,22	115,96	110,13
Accuracy (%)	100	100	100

- Scenario 2

**Table 3.** Scenario 2

Output	Boyer Moore	Horspool	Zhu Takoka
Runtime (s)	0,118	0,086	0,052
Memory Consumption (kb)	110,19	115,95	110,01
Accuracy (%)	100	100	100

- Scenario 3

**Table 4.** Scenario 3

Output	Boyer Moore	Horspool l	Zhu Takoka
Runtime (s)	0,114	0,098	0,12
Memor Consumption (kb)	110,14	115,91	110,05
Accuracy (%)	100	100	100

- Scenario 4

**Table 5.** Scenario 4

Output	Boyer Moore	Horspool l	Zhu Takoka
Runtime (s)	0,112	0,106	0,108
Memory Consumption (kb)	110,13	115,98	110,04
Accuracy (%)	100	100	100

- Scenario 5

**Table 6.** Scenario 5

Output	Boyer Moore	Horspool l	Zhu Takoka
Runtime (s)	0,106	0,106	0,084
Memory	110,13	115,98	110,04
Consumption (kb)			
Accuracy (%)	100	100	100

- Scenario 6

**Table 7.** Scenario 6

Output	Boyer Moore	Horspool l	Zhu Takaoka
Runtime (s)	0,094	0,114	0,136
Memory	110,05	115,8	109,97
Consumption (kb)			
Accuracy (%)	100	100	100

- Scenario 7

**Table 8.** Scenario 7

Output	Boyer Moore	Horspool	Zhu Takoka
Runtime (s)	0,09	0,118	0,11
Memory	110,02	115,78	109,94
Consumption (kb)			
Accuracy (%)	100	100	100

- Scenario 8

**Table 9.** Scenario 8

Output	Boyer Moore	Horspool	Zhu Takoka
Runtime (s)	0,072	0,12	0,104
Memory	109,98	115,74	109,89
Consumption (kb)			
Accuracy (%)	100	100	100

### 3.8 Output Analysis

In this last stage, the researcher analyzed the outputs and simulations carried out at the time of experimentation. The output is represented in the form of tables and graphs that state the value of processing time (runtime), the memory used (memory consumption), and the accuracy (accuracy) of each algorithm at the time of each scenario. The final discussion will discuss the overall analysis of the results of all simulation scenarios. In addition, the analysis discusses the relationship between input data values and the influence of each variable. The output of the scenario will be calculated with the Exponential Comparison Method.

- The process of calculating the total value in the 1st process:

$$\begin{aligned}\text{Boyer Moore Value} &= (0,158)^{0,5} + (110,22)^{0,5} \\ &= 0,397492 + 10,49857 \\ &= 11,0583\end{aligned}$$

$$\begin{aligned}\text{Horspool Value} &= (0,084)^{0,5} + (115,96)^{0,5} \\ &= 0,289828 + 10,768473 \\ &= 10,89606\end{aligned}$$

$$\begin{aligned}\text{Zhu Takaoka Value} &= (0,138)^{0,5} + (110,13)^{0,5} \\ &= 0,371484 + 10,49428 \\ &= 10,86577\end{aligned}$$

- The process of calculating the total value in the 2nd process:

$$\begin{aligned}\text{Boyer Moore Value} &= (0,118)^{0,5} + (110,19)^{0,5} \\ &= 0,343511 + 10,49714 \\ &= 10,84065\end{aligned}$$

$$\begin{aligned}\text{Horspool Value} &= (0,086)^{0,5} + (115,95)^{0,5} \\ &= 0,293258 + 10,768008 \\ &= 11,06127\end{aligned}$$

$$\begin{aligned}\text{Zhu Takaoka Value} &= (0,052)^{0,5} + (110,01)^{0,5} \\ &= 0,228035 + 10,48857 \\ &= 10,7166\end{aligned}$$

- The process of calculating the total value in the 3th process:

$$\begin{aligned}\text{Boyer Moore Value} &= (0,114)^{0,5} + (110,14)^{0,5} \\ &= 0,337639 + 10,49476 \\ &= 10,8324\end{aligned}$$

$$\begin{aligned}\text{Horspool Value} &= (0,098)^{0,5} + (115,91)^{0,5} \\ &= 0,31305 + 10,766151 \\ &= 11,0792\end{aligned}$$

$$\begin{aligned}\text{Zhu Takaoka Value} &= (0,12)^{0,5} + (110,05)^{0,5} \\ &= 0,34641 + 10,49047 \\ &= 10,83688\end{aligned}$$

- The process of calculating the total value in the 4th process:

$$\begin{aligned}\text{Boyer Moore Value} &= (0,112)^{0,5} + (110,13)^{0,5} \\ &= 0,334664 + 10,49428 \\ &= 10,82895\end{aligned}$$

$$\begin{aligned}\text{Horspool Value} &= (0,106)^{0,5} + (115,98)^{0,5} \\ &= 0,325576 + 10,769401 \\ &= 11,09498\end{aligned}$$

$$\begin{aligned} \text{Zhu Takaoka Value} &= (0,108)^{0,5} + (110,04)^{0,5} \\ &= 0,328634 + 10,49 \\ &= 10,81863 \end{aligned}$$

- The process of calculating the total value in the 5th process:

$$\begin{aligned} \text{Boyer Moore Value} &= (0,106)^{0,5} + (110,06)^{0,5} \\ &= 0,325576 + 10,49095 \\ &= 10,81652 \end{aligned}$$

$$\begin{aligned} \text{Horspool Value} &= (0,106)^{0,5} + (115,84)^{0,5} \\ &= 0,325576 + 10,762899 \\ &= 11,08848 \end{aligned}$$

$$\begin{aligned} \text{Zhu Takaoka Value} &= (0,084)^{0,5} + (109,97)^{0,5} \\ &= 0,289828 + 10,48666 \\ &= 10,77649 \end{aligned}$$

- The process of calculating the total value in the 6th process::

$$\begin{aligned} \text{Boyer Moore Value} &= (0,094)^{0,5} + (110,05)^{0,5} \\ &= 0,306594 + 10,49047 \\ &= 10,79707 \end{aligned}$$

$$\begin{aligned} \text{Horspool Value} &= (0,114)^{0,5} + (115,8)^{0,5} \\ &= 0,337639 + 10,761041 \\ &= 11,09868 \end{aligned}$$

$$\begin{aligned} \text{Zhu Takaoka Value} &= (0,136)^{0,5} + (109,97)^{0,5} \\ &= 0,368782 + 10,48666 \\ &= 10,85544 \end{aligned}$$

- The process of calculating the total value in the 7th process:

$$\begin{aligned} \text{Boyer Moore Value} &= (0,09)^{0,5} + (110,02)^{0,5} \\ &= 0,3 + 10,48904 \\ &= 10,78904 \end{aligned}$$

$$\begin{aligned} \text{Horspool Value} &= (0,118)^{0,5} + (115,78)^{0,5} \\ &= 0,343511 + 10,760112 \\ &= 11,10362 \end{aligned}$$

$$\begin{aligned} \text{Zhu Takaoka Value} &= (0,11)^{0,5} + (109,94)^{0,5} \\ &= 0,331662 + 10,48523 \\ &= 10,81689 \end{aligned}$$

- The process of calculating the total value in the 8th process:

$$\begin{aligned} \text{Boyer Moore Value} &= (0,072)^{0,5} + (109,98)^{0,5} \\ &= 0,268328 + 10,48713 \\ &= 10,75546 \end{aligned}$$



$$\begin{aligned} \text{Horspool Value} &= (0,12)^{0,5} + (115,74)^{0,5} \\ &= 0,34641 + 10,758253 \\ &= 11,10466 \end{aligned}$$

$$\begin{aligned} \text{Zhu Takaoka Value} &= (0,104)^{0,5} + (109,98)^{0,5} \\ &= 0,32249 + 10,48284 \\ &= 10,80533 \end{aligned}$$

- Calculate the priority value of the decision

$$\begin{aligned} \text{Total Value Boyer Moore} &= 10,89606 + 10,84065 + 10,8324 + 10,82895 + \\ &10,81652 + 10,79707 + 10,78904 + 10,75546 \\ &= 86,55015 \end{aligned}$$

$$\begin{aligned} \text{Total Value Horspool} &= 11,0583 + 11,06127 + 11,0792 + 11,09498 + 11,08848 + \\ &11,09868 + 11,10362 + 11,10466 \\ &= 88,68919 \end{aligned}$$

$$\begin{aligned} \text{Total Nilai Zhu Takaoka} &= 10,86577 + 10,7166 + 10,83688 + 10,81863 + 11,08848 \\ &+ 10,85544 + 10,81689 + 10,80533 \\ &= 86,80402 \end{aligned}$$

- Determine Results or Priority Decisions

After obtaining the final value or total value of each, alternative, the next step that needs to be done is to determine the priority of the decision based on the value of each alternative. The results of priority decisions can be seen in the table below:

Alternative	Total	Ranking
Boyer Moore Algorithm	86,55015	1
Zhu Takaoka Algorithm	86,80402	2
Horspool Algorithm	88,68919	3

From the results of calculations using the Exponential Comparison Method, it can be seen that the most effective algorithm with runtime parameters and memory consumption is the Boyer Moore Algorithm. Next, the second position is the Zhu Takaoka Algorithm and the last position is the Horspool Algorithm. However, each of these algorithms has their advantages and disadvantages. From the performance analysis that the author has done, for runtime parameters, Boyer Moore Algorithm is the best algorithm used for word search with a large number of characters, while Horspool algorithm is the best algorithm used if you want to find words with a small number of characters.

For memory consumption parameters, the more characters are searched, then each algorithm will need less memory. However, from the results of the analysis that the author has done, the smallest memory consumption is the Zhu Takaoka algorithm and is followed by the Boyer Moore algorithm.

## 4 Conclusion

The results of the comparison of the Boyer Moore Algorithm, Horspool Algorithm, and Zhu Takaoka Algorithm in the Indonesian hadith translation repository using a simulation method consisting of problem formulation stages, conceptual models, input and output data, modeling, simulation, verification and validation, experimentation, and output analysis and after calculating with the Exponential Comparison Method (MPE) with runtime and memory consumption parameters shows that the Boyer Moore Algorithm has the best value, followed by the Horspool Algorithm and finally the Zhu Takaoka Algorithm.

## 5 Sugestion

1. The algorithm used or compared can be added with other String Matching algorithms, such us Turbo-Boyer Moore or Tuned-Boyer Moore
2. To find out the broader comparative results can use a programming language different from the implementation to other objects than the author uses.

**Acknowledgements.** This paper in conjunction with The 1st International Conference on Islam, Science and Technology (ICONIST 2018) in Malang, East Java.

## References

- [1] Sitorus, Lamhot. 2015. Algoritma dan Pemrograman. Yogyakarta : CV Andi Publ.
- [2] Togatorop. 2017. Implementasi Algoritma Zhu Takaoka. Vol. 5(2), 147-153
- [3] Tambun. 2017. Perbandingan Penggunaan Algoritma BM dan Algoritma Horspool pada Pencarian String dalam Bahasa Medis
- [4] Hansun, dkk. 2014. Implementasi Algoritma Boyer Moore pada Aplikasi Kamus Kedokteran Berbasis Android. Tangerang.
- [5] Geraldy, Ditya, dkk. 2016. Studi Perbandingan Algoritma Brute Force, Algoritma Knuth Morris Pratt- Algoritma Boyer Moore untuk Identifikasi Kesalahan Penulisan Teks Berbasis Android. Salatiga.
- [6] Fau, dkk. 2017. Analisa Perbandingan Boyer Moore dan Knuth Morris Pratt dalam Pencarian Judul Buku Menerapkan Metode Perbandingan Eksponensial. Medan.
- [7] Kusnadi, Adhi, dkk. 2017. Perbandingan Algoritma Horspool dan Algoritma Zhu Takaoka dalam Pencarian String Berbasis Dekstop. Serpong. UMN.
- [8] Handrizal, dkk.2017. analysis of the Zhu Takaoka Algorithm and the Knuth Morris Pratt Algorithm that is implemented in the dictionary based on android with time as a parameter.
- [9] Aulia, A., Khairani, D., Hakiem N., and Bahawares, R. B. (2017). WatsaQ : Repository of Al hadith in Bahasa (Case Study : Hadith Bukhori). 4th International Conference on Electrical Engineering Computer Science and Informatics (EECSI).
- [10] Fadly Fachri, Masruroh, S. U., and Suseno H. B (2017). Aes-Gcm And Aes-Eax Evaluation Of Aes-Gcm And Aes-Eax Connected Encryption Performance.
- [11] Aulia, A., Khairani, D., Hakiem N. (2017). Development of a Retrieval System for Al Hadith in Bahasa (Case Study : Hadith Bukhari). Conference on Cyber and IT Service Management.