
Android-Based Horspool Algorithm for Proverb Search

Andri Agus Yahya

Stmik Budi Darma

Email : andriagus@gmail.com

Received : 20 April 2021

Revised : 19 Mei 2021

Accepted : 21 Juny 2021

© 2021 The Author: Published by. Cattleya Darmaya Fortuna

Abstrak

In the context of linguistics, a proverb is a phrase or a group of words that expresses an intention or will, a person's state of mind, or anything that says behavior, actions, and other things about someone, and which indirectly reflects something that the reader or listener may comprehend. Because not everyone understands proverbs, an android-based proverb search application makes it simpler for someone to search for them, and it is more practical and efficient than a web-based search engine. Following this investigation, the outcomes of character matching using the Horspool algorithm were finished in the sixth iteration, according to the results of the study.

Keywords: Apps, Proverbs, Horspool Algorithm, Android.

INTRODUCTION

Proverbs have existed since before Malays were capable of writing and reading. The Malays had been educated through proverbs before they learned to write and read. A proverb is a sentence or a set of words that reflect a purpose or will, a person's circumstances, or something that expresses one's conduct, activities, and personal characteristics. The proverb must be conveyed in an engaging manner [1]. Proverbs can also be interpreted as an expression that, while still not directly communicating something to the reader or listener, becomes to communicate successfully. The process of conveying information is conveyed by written or oral means to make it easier for someone to convey the intent of thoughts, ideas, and decisions in verbal communication[2]. Speakers who use an attractive linguistic style and aphorisms will better engage listeners and influence them. [3]. The primary objective of proverbs is to serve as a form of communication, teaching, and satire using subtle phrases so that those who are scolded are not easily offended, and this applies to both older and younger people. For example, if someone wants to criticize the mentality of someone older than him, he can use a proverb with subtle humor or a sarcastic remark. [4].

The search for a pattern in a text is known as string matching. String matching is a technique for locating a pattern in a text string. The heuristic search method is a way of finding the target state that is based on a set of rules [5]. Because the spelling is incorrect, typing errors lead the standard term to become a non-standard word. This can be remedied by creating a system that detects typing errors [6]. The distance between two strings is calculated using these procedures. The closer the two strings are, the closer they are

considered to match [7]. It is possible to utilise the journal search mobile application by searching for terms that are required by Android-based smartphone users. [8]. Matching Characters is a kind of method that attempts to match characters in a collection of text; this method works to match a character known as a pattern in a collection of text; String Matching is one such type of method [9].

Horspool's method employs terms in the search process as patterns, as well as vocabulary that has been put into the database as text, to get the best results [10]. Search efficiency in translating the Bima language into Indonesian is accomplished via the use of a string matching algorithm in Bima – Indonesian language dictionaries depending on the purpose of the author. The Horspool method was used in this research to find matches between strings of characters [11]. Although word matching is done from right to left in Horspool's method, the string search time is reduced because stored search information is used to execute subsequent shifts [12]. It was R. Nigel Horspool who came up with a simplified version of the Boyer-Moore algorithm, which he named after himself. The issue with this search query is that it requires sifting through a large amount of information in order to discover the first pattern. So, because the text being searched for may be very big (hundreds of thousands or even hundreds of lines), it is critical to use methods that are more efficient. In contrast to the Boyer-Moore algorithm, the Horspool algorithm employs a technique that is almost similar to the Boyer-Moore algorithm, but it does not leap depending just on letters in the pattern that are determined to be inappropriate in the text. Using Horspool's method, you may move the right-hand letter from the window by a certain amount. The shift value for each character will be computed at the beginning of the observation stage (preprocessing) of the procedure. It is at this point that the pattern is examined from side to side till a match or a pattern mismatch is discovered. When shifting values, every right-hand letter in the pane is used as an index to locate the new values. In the event of a mismatch (a character does not appear in the pattern), the window is moved by the length of the pattern that caused the mismatch. Instead, the window is moved in accordance with the character to the right of the pattern's first character.

METHODS

Android Application Design

People employed in this area process business information or technical information that is used to assist business operations or technical management/decision-making. Application software, in addition to the usual data processing applications, is used to manage the real operations of businesses. In this instance, an android-based dictionary application is one of the android-based apps that may make it simpler for users to find what they are looking for in a word search application. A research was carried out to determine the community's requirements for utilising a dictionary in the android-based word search process, as well as for designing and developing computer-assisted apps, which included display design and manufacturing using the Eclipse Mars programming language [13]. Android is created utilising the Android programming language and the SQLite database that it relies on for its backend [14].

Proverb

These are phrases or sets of words that convey a purpose or will, the condition of a person, or anything that characterises a person's conduct, activities, and other characteristics. Proverbs may also be regarded as a phrase that communicates something to the reader or listener, although not directly, but implicitly, without the reader or listener having to think about it. In order to serve as linguistic guidance without the need to carry a dictionary in the form of a book with a thickness and weight that is unusually hefty for the size of a book, proverbs must be communicated in an appealing style and manner. The primary aim of the development of proverbs is to serve as a method of communication, teaching, and satire via the use of delicate language, so that those who are admonished are not easily offended,

regardless of whether they are reprimanding those who are older or younger in age. To criticise the attitude of someone who is older than him, for example, someone may use a proverb that includes subtle sarcasm or counsel about the attitude of the person who is being scolded.

Search With Boyer-Moore Horspool Algorithm Proverb

The Horspool algorithm employs a technique that is almost identical to the Boyer-Moore algorithm, with the exception that it does not leap depending on the characters in the pattern that are determined to be inappropriate for use in the text. [15]. Using Horspool's method, you may move the rightmost character from the window by a certain amount. The shift value for each character will be computed at the first observation step (preprocessing) [16]. After that, the pattern is examined from right to left until a match or design mismatch is identified. When utilising the Horspool algorithm, there are two phases to the matching of characters to strings, which are as follows:

1. Preprocess stage

At this point, pattern observations are taken in order to construct a bad-match table, which includes shift values in the event of a mismatch between the pattern and the text. The following are the stages that the horse pool algorithm does in a systematic manner during the preparation stage:

- a. Each character in the pattern is added to the bad-match database, and the shift value is computed for each character in the pattern.
- b. Characters that are at the end of the pattern are not counted and are not converted into the right-hand character of the same character as the character that they are at the end of the pattern.
- c. If there are two characters that are the same and one of them is not the right-hand character, then the shift value is computed using the character with the greatest index, rather than the character with the smallest index.
- d. When a character in the text is not found in the pattern, Horspool's method defaults to storing the length of the pattern as the length of the shift value, which is a good thing.
- e. May get the shift value to use by dividing the length of the pattern by the last index of the character minus one for each character, yielding value = $m-1$ for each character.

2. Search Stage

The following are the stages taken by the Horspool algorithm during the preparation stage, in a systematic manner:

- a. A comparison between pattern's rightmost character and its corresponding character in the window
- b. When there is a mismatch between two characters, the bad-match table is utilised to skip characters.
- c. When there is a mismatch, the rightmost character in the window is used to calculate the distance between the two characters that must be shifted.
- d. Following a successful match, the window is moved to the right (whether the results match or do not match).
- e. This process is continued until the text in the window is completed, or until the pattern matches the text in the window.

RESULTS & DISCUSSION

Indonesian Proverb Search Process Display

When looking for Indonesian proverbs, the layout display of the Indonesian proverb search process is shown on the computer screen.



Figure 1: Indonesian Proverb Search

Display the Meaning of Indonesian Proverbs

The Layout view of the meaning of proverbs is a view that displays the meaning of the proverbs that have been searched for and have been previously chosen to be shown. The layout comprises of a text view that displays the meaning of the proverbs that have been selected.



Figure 2 : Meaning of Indonesian Proverbs

The picture below depicts a search result obtained via string matching, and the view in this

view represents the outcome of the implementation in the language search process.



Figure 3 : Language Search Results

As may be seen in the preceding table, the initialization of the initials of the bad-match has been accomplished. It is given the numbers m and I for each line of text and pattern, where m represents both the length of each design and the index of the design.

Table 1 : The First Iteration of the Boyer-Moore Horspool Algorithm

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
P	B	A	G	A	I	_	A	I	R	_	D	I	_	D	A	U	N	_	T	A	L	A	S
T	T	A	L	A	S																		
1	0	1	2	3	4																		

In the table, there is a mismatch between the characters "I" and "S" in the rightmost window of the text, which is highlighted in red. When creating the bad-match table, the "I" character has a shift value of five, which is because the "I" character was unknown at the time of creation of the table. As a result, make five shifts to the right on the window.

Table 2: Second Iteration of the Boyer-Moore Horspool Algorithm

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
P	B	A	G	A	I	_	A	I	R	_	D	I	_	D	A	U	N	_	T	A	L	A	S
T							T	A	L	A	S												
1						0	1	2	3	4													

Another mismatch exists between the characters "_" and "S" in the text's rightmost window, which is caused by the letter "S." The "_" character has a shift value of five in the bad-match table, which is due to the fact that the "_" character was unknown when the bad-match database was created. As a result, make five shifts to the right on the window.

Table 3: Third Iteration of the Boyer-Moore Horspool Algorithm

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
P	B	A	G	A	I	_	A	I	R	_	D	I	_	D	A	U	N	_	T	A	L	A	S
T											T	A	L	A	S								
1											0	1	2	3	4								

Another mismatch exists between the letters "A" and "S" in the text's rightmost window, which is caused by the letter "A." Due to the fact that the character "A" is a known character in the process of constructing the bad-match table, the shift value of the character "A" is one in the bad-match table. As a result, move the cursor one time to the right on the window.

Table 4: Fourth Iteration of Boyer-Moore Horspool Algorithm.

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
P	B	A	G	A	I	_	A	I	R	_	D	I	_	D	A	U	N	_	T	A	L	A	S
T												T	A	L	A	S							
1												0	1	2	3	4							

A mismatch exists between the characters "U" and "S" in the text's rightmost window, which is shown by the letter "U." The sliding value of the character "U" in bad-match tables is five, since the sign "U" was not used in the creation of bad-match tables, resulting in a five-point penalty. Consequently, you may move to the right of the window as many times as you like.

Table 5: Fifth Iteration of the Boyer-Moore Horspool Algorithm.

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
P	B	A	G	A	I	_	A	I	R	_	D	I	_	D	A	U	N	_	T	A	L	A	S
T																	T	A	L	A	S		
1																	0	1	2	3	4		

Another mismatch occurs between the letters "L" and "S" in the text's rightmost window, which occurs between the letters "L" and "S." When creating a bad-match table, the "L" character has a shift value of two because the letter "L" is a commonly used character in the creation of bad-match tables. As a result, make two shifts to the right on the window.

Table 6: Sixth Iteration of the Boyer-Moore Horspool Algorithm.

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
P	B	A	G	A	I	_	A	I	R	_	D	I	_	D	A	U	N	_	T	A	L	A	S
T																			T	A	L	A	S
1																			0	1	2	3	4

The window is located at the conclusion of the text, and all of the patterns correspond to the text. It has been determined that all character matching using the Horspool algorithm has been accomplished, and that it has reached the sixth iteration. There was a successful search for the specified term.

CONCLUSION

The following are some conclusions drawn from the issues raised above: Due to the limits of the smartphone screen, the process of searching for Indonesian proverbs is required, and the Indonesian proverbs are highly sought for, making the process of looking for Indonesian proverbs necessary. The Boyer Moore Horspool Algorithm may be used to search for Indonesian proverbs, which is a useful tool. The Eclipse 4.2 Juno application is compatible with Android versions 2.2 Gingerbread and above, all the way up to Android 5.1 Lollipop.

REFERENCES

- [1] A. Irawan, "Pembuatan Aplikasi Game Kuis Tebak Gambar Tau Ke Pontianak Berbasis Android," *J. Khatulistiwa Inform.*, vol. 5, no. 2, 2017.

- [2] I. H. Khotimah, "Komunikasi verbal dan non verbal dalam diklat," *Komun. Verbal dan Non Verbal Dalam Diklat*, no. November, 2019.
- [3] Y. Wardhani, "Pemakaian Pribahasa dan Kata Mutiara dalam Retorika Dakwah Para Da'i di Indonesia : Kajian Stalistika dalam Sastra Arab-Indonesia," *Bul. Al-Turas*, vol. 24, no. 2, 2018, doi: 10.15408/bat.v24i2.8667.
- [4] Haryati, "THE ANALYSIS OF ELLIPTICAL NOUN PHRASES OF THE BOOK A COMPLETE COLLECTION OF ENGLISH PROVERBS BY USING TRANSFORMATIONAL GENERATIVE GRAMMAR AT ENGLISH DEPARTMENT, ANDALAS UNIVERSITY," *Vivid J.*, vol. 44, no. 11, 2015.
- [5] S. Suryadi, "PERANCANGAN APLIKASI PENCARIAN FILE DENGAN MENGGUNAKAN METODE BEST FIRST SEARCH," *J. Inform.*, vol. 2, no. 2, 2019, doi: 10.36987/informatika.v2i2.197.
- [6] Y. Rochmawati and R. Kusumaningrum, "Studi Perbandingan Algoritma Pencarian String dalam Metode Approximate String Matching untuk Identifikasi Kesalahan Pengetikan Teks," *J. Buana Inform.*, vol. 7, no. 2, 2016, doi: 10.24002/jbi.v7i2.491.
- [7] M. O. Braddley, M. Fachrurrozi, and Y. Novi, "Pengoreksian Ejaan Kata Berbahasa Indonesia Menggunakan Algoritma Levensthein Distance," *Pros. Annu. Res. Semin.*, vol. 3, no. 1, 2017.
- [8] B. H. Lukmana, "Penerapan Algoritma Crochemore-Perri Pada Pencarian Jurnal Berbasis Mobile," *J. Pelita Inform.*, vol. 17, 2018.
- [9] F. Ikhwan, "IMPLEMENTASI ALGORITMA HORSPPOOL PADA APLIKASI TAKSONOMI BERBASIS ANDROID," *Buffer Inform.*, vol. 4, no. 2, 2018, doi: 10.25134/buffer.v4i2.1474.
- [10] V. Frigustini, A. Erlansari, and D. Andreswari, "Implementasi Algoritma Horspool Pada Aplikasi Kamus Bahasa Lintang - Indonesia Berbasis Android," *Rekursif*, vol. 6, no. 1, 2018.
- [11] S. Bill, P. Ginting, N. Marbun, M. Zarlis, and D. Hartama, "Penerapan Algoritma Horspool Perancangan Aplikasi Kamus Bahasa Bima-Indonesia," *Semin. Nas. Mat. dan Terap.*, vol. 1, 2019.
- [12] D. Kartini, R. Y. Tanjung, and I. Lasmana, "Penerapan Algoritma Boyer Moore Horspool Pada Pencarian Katalog Buku," 2019.
- [13] Z. Abidin, A. P. Sari, and R. Nasution, "Aplikasi Pencarian Kata dalam Bahasa Sunda Menggunakan Algoritma Boyer Moore Berbasis Android," *Konf. Nas. ilmu Sos. Teknol.*, 2015.
- [14] B. D. Indiana and I. Ramadhani, "APLIKASI PEMBELAJARAN BAHASA JAWA BERBASIS ANDROID," *CAHAYAtech*, vol. 8, no. 1, 2019, doi: 10.47047/ct.v8i1.18.
- [15] T. S. Alasi, "Algoritma Boyer Moore Untuk Penyaringan Pesan Teks Menggunakan Perbandingan Kata Yang Sama," *Semin. Nas. Sains Teknol. Inf.*, 2018.
- [16] A. Kusnadi and A. K. Wicaksono, "Perbandingan Algoritma Horspool dan Algoritma Zhu-Takaoka dalam Pencarian String Berbasis Desktop," *J. Ultim. Comput.*, vol. 9, no. 1, 2017, doi: 10.31937/sk.v9i1.568.