Visual Approach of Searching Process using Boyer-Moore Algorithm

Robbi Rahim1*, Ansari Saleh Ahmar2, Ayu Putri Ardyanti3 and Dicky Nofriansyah4

1Department of Health Information, Akademi Perekam Medik dan Infokes Imelda, Jl. Bilal Ujung Medan 20116, Indonesia
2Postgraduate Ph.D. Student Universiti Malaysia Perlis, Universiti Malaysia Perlis, Malaysia
3Department of Statistics, Universitas Negeri Makassar, Daeng Tata Street, UNM Campus, Makassar City, Indonesia
4Department of Informatics Engineering, STMIK Primakara, Jalan Tukad Badung No.135 Denpasar Bali, Indonesia
5Department of Information System, STMIK Triguna Dharma, Jl. Jenderal Abdul Haris Nasution No.73, Medan 20219, Indonesia

*usurobbi85@zoho.com

Abstract. This research shows the process of string matching using a Boyer-Moore algorithm in the form of visual simulation by doing string searching of a pattern string, the result of visualization is recognized that Boyer-Moore algorithm will compare a n string from the right and will test to n-1 String to get the appropriate character of the desired input string.

1. Introduction

Today's digital era is very easy to get information, this is because the higher the advancement of technology such as the internet, smartphones that can be used by many users, to get information users can use search engines like Yahoo, Google or other search engines [1] [2], one of the techniques that can be used to search is the Boyer-Moore algorithm.

The Boyer-Moore algorithm has a string matching process from the right of the pattern [3] [4] [5], the initial concept of the string-handling process from the right to get more information [3] [5], this research tries to model in the simulation of string matching process using Boyer-Moore algorithm, it is expected for other researchers who wish to study or developed the Boyer-Moore algorithm to find out how the algorithm works.

2. Theory

Searching is a work that is usually performed in everyday life, the search process is done a job to find a word or replace a particular word [6] [7], and Boyer-Moore algorithm is one algorithm that could use for string matching.

Boyer Moore's algorithm was published by Robert S. Boyer and J. Strother Moore in 1977. The character matching made using the Boyer-Moore algorithm starts from the right-hand strings of the
pattern. The idea behind this algorithm is to start matching characters from the right, and then more information will be obtained.

The Boyer-Moore algorithm includes the most efficient string matching algorithm compared to other string matching algorithms due to an efficient nature of the algorithm, many string matching algorithms developed based on the concept of Boyer-Moore algorithm, some of which are the Turbo BM algorithm and Quick Search algorithm, as for the string search steps on the Boyer-Moore algorithm as follows:

a. First, we need two tables with Match Heuristic (MH), and Occurrence Heuristic (OH) approaches to determine the number of shifts that will be performed on a pattern (P) if there are unsuitable characters in the matching process of characters in text (S).
b. If in the comparison process there is character mismatch between the characters on P and S, then the shift is done by looking at both tables with the largest shift value selected.
c. The possibility of completion in shifting to P is that if in previous matching there is no matching character then the shift is done by looking at the value of shift in occurrence heuristic table. If the character being compared does not exist in P, then the shift is done as much as the number of characters contained in P, but if the unsuitable character is contained in the P string, then the shift is done based on the table.
d. If the characters in the matching text match the characters on the P string, then the character checking position on P and S shifts each left position 1 from the previous position, then proceed with matching at that position and so on, then if there is a character mismatch In P and S, the shift is done by looking at the heuristic match table and the occurrence heuristic where the largest shift value to be selected is reduced by the number of matching characters.
e. If all the characters have a match, that means P has been found in S, then move the pattern by one character, continue until the end of the S pattern.

3. Result and Discussion
Here is an example of Boyer-Moore algorithm analysis for word search process, it is known that the input string is DAELYMMAKRINAAMRSQ and the target string is RINA, and here is the process of applying the Boyer-Moore algorithm to find the target string from the string of entry.

a. Declare input strings and target strings

| DAELYMM | MAKRINA | AMRS | W |
|---------|---------|-----|---|
| R | I | N | A |

b. First Step
The last character of the target string 'A' does not match the 'L' character of the word 'DAELYMMAKRINAAMRSQ', the first checking process is done by counting the length of the target string n and then the last letter of the target string is matched to the nth character of the input string.

| DAELYMM | MAKRINA | AMRS | W |
|---------|---------|-----|---|
| R | I | N | A |

Seen in the process above the letter A != L then the next process is to test L with all the letters that exist in the target string as below
Since nothing matches then the target string shifts to the right as much as the length of the input string

c. Second Step

It appears in the process above the letter A == A then the next process is to test all the characters contained in the target string as much as n the target string of the input string, the following results

It appears in the above process that the corresponding characters are only characters A and not for other characters then the next process of the target string goes forward as many as the target string n characters

d. Third Step

Seen in the process above the letter A != N then the next process is to test all the characters contained in the target string as much as n the target string of the input string, the following results.

in the above process there are some similar characters between the target strings with the string of entry, based on the function of the Boyer-Moore algorithm if there is the same character then the character position between the target string and the goal string must be equated as below
Seen in the above process the corresponding character position is the same between the target string position and the input string, then the above process is checked again to test whether the target string character is the same as the character of the input string.

In the above process, it appears that the last character of the target string ‘A’ is the same as the ‘A’ input string for the same position, after which it tested all the same character positions between the target string and the input string.

Appear in the table above is appropriate then the process is stopped, and the word "RINA" is located in characters 10 to 13, then the remaining characters of the input string are not processed.

4. Conclusion

The process visualization of the Boyer-Moore algorithm allows researchers or scholars to learn how the Boyer-Moore algorithm works, and in the application development it will be easier to create a function for word search and can be implemented into many search processes.

References

[1] N. R. Dalal and P. Jadhav, "A Composite Algorithm for String Matching," *International Journal of Modern Trends in Engineering and Research (IJMTER)*, vol. 2, no. 7, pp. 68-73, 2015.

[2] R. Janani and S. Vijayarani, "An Efficient Text Pattern Matching Algorithm for Retrieving Information from Desktop," *Indian Journal of Science and Technology*, vol. 9, no. 43, pp. 1-11, 2016.

[3] A. Z. M. Saleh, N. A. Rozali, A. G. Buja, K. A. Jalil, F. H. Mohd Ali and T. F. A. Rahman, "A Method for Web Application Vulnerabilities Detection by Using Boyer-Moore String Matching Algorithm," in *The Third Information Systems International Conference*, 2015.

[4] A. A. Alsaheel, A. H. Alqahtani and A. M. Alabdulatif, "Analysis of Parallel Boyer-Moore String Search Algorithm," *Global Journal of Computer Science and Technology Hardware & Computation*, vol. 13, no. 1, pp. 43-47, 2013.

[5] R. Y. Tsarev, A. S. Chernigovskiy, E. A. Tsareva, V. V. Brezitskay, A. Y. Nikiforov and N. A. Smirnov, "Combined string searching algorithm based on Knuth-Morris-pratt and Boyer-Moore algorithms," in *XIX International Scientific Conference Reshetnev Readings*, Rusia, 2015.
[6] G. Pandey and G. L. Prajapati, "Applying Bi-Directional Search Strategy in Selected String Matching Algorithms," *International Journal of Computer Applications*, vol. 143, no. 11, pp. 40-43, 2016.

[7] C. S. Rao, K. B. Raju, and S. V. Raju, "Parallel String Matching with Multi-Core Processors - A Comparative Study for Gene Sequences," *Global Journal of Computer Science and Technology Hardware & Computation*, vol. 13, no. 1, pp. 26-41, 2013.