



Application Of Boyer Moore Algorithm for Text Searching

Fince Tinus Waruwu

Department of Computer Science STMIK Budi Darma, North Sumatera, Indonesia

Abstract – Technology at the moment is very rapid development in business, and education. Many developers or application makers are competing in improving science about the science of technology so as not to miss. Technology is an integral part of human life. Computer application is one part that is not separated from the computer. An application is a software that actually works to simplify the job. Thus, at this time already many computer applications are created. In the course, the application software always experienced improvements made by the manufacturer. These improvements lead to improved facilities to make it easier for users to work. In the world of education or the world of work a lot of work done by using computers such as typing by using text editor applications such as Microsoft Word in the manufacture of scientific work or other work, often mistakes in terms of writing words, often mistakes occur not because of intentional but because of community habits In abbreviating words when writing. It is necessary an application that can facilitate the examination of a writing in order to find the wrong word or abbreviated and automatically the application will fix it (text improvement). To make it easier to find the wrong word in the application, it needs string matching technique using Boyer Moore algorithm. The application is built with Microsoft Visual Studio 2008 and MySQL as the.

Keywords - String Matching, Search Pattern, Text, Boyer Moore

1 INTRODUCTION

Computer application is one part that is not separated from the computer. The application software actually works to simplify the job. Thus, at this time already many computer applications are created. In the current development of software applications always experience improvements made by the manufacturer. These improvements lead to improved facilities to make it easier for users to work. The text editor is one of the existing software in computer technology. The text editor is one of the integral tools in human life, especially in the field of education and the world of work. The text editor is a text processing application that serves to create or edit text such as notepad, WordPad, Microsoft Word and much more other types of application editor.

In the world of education or the world of work a lot of work done by using computers such as typing by utilizing a text editor software such as Microsoft Word in the manufacture of scientific work or other work, Often mistakes in terms of writing words, often mistakes happen not because of intentional but because of community habits In abbreviating words when doing typing. So in the examination of a writing to find the wrong word required accuracy in reading and requires time not too little, due to time constraints and the number of human work to be completed so that the need to develop application technology that can find the wrong text and then repaired automatically.

Based on the above problems it is necessary an application that can facilitate the examination of a writing in order to find the wrong word or abbreviated and automatically the application will be repaired. The text repair application that will be designed requires a text file as the source of the text to be checked. A text file is a file containing information in the form of text or data derived from a word processing application. So with the application of course will be very helpful for people who want to check posts quickly. So, people no longer need to check or observe the writing manually that can take time is not small. To make it easier to find the wrong word in the application, it needs a searching technique (string matching). String matching or often called string matching is a technique for searching all occurrences of a short string pattern $[0..n-1]$ called pattern in a longer string of text $[0..m-1]$ called text. One of the known string matching algorithms is Boyer Moore. The text-fixing app will be designed using the Boyer Moore algorithm. Boyer Moore string matching algorithm is one of the famous algorithms used in string search algorithms [1]. Based on the searching direction, Boyer Moore's algorithm is classified as an algorithm that reads strings from left to right. Therefore, Boyer Moore's algorithm has the most natural string reading direction.

2 THEORY

2.1 Application



The application is the act of acquiring and integrating the physical and conceptual resources that result in a system [2]. Steps in the implementation stage are:

1. Designing implementation. Since only one stage of development is left before the new system is used the manager and information specialist understand well the work required to implement the system.
2. Announce implementation. The implementation project is announced to employees in the same manner as in system research. The purpose of this announcement is to inform employees about the decision to implement a new system and to ask for employee cooperation.

Get hardware resources. System design is provided for suppliers of various types of computer equipment contained in approved configurations. The information specialist provides support for this decision by studying the proposal and making recommendations.

2.2 Algorithm

The algorithm is the logic, method and systematic (sequence) step used to solve a problem [3]. In some contexts, the algorithm is a sequence specification step to do a particular job. Considerations in the selection of algorithms are to be true, caring, and efficient. The term algorithm comes from the name of an Arab author named Ja'fat Mohammed bin Musa al Khowarizmi (790 - 840), who is very famous as the father of Algebra. By definition, the algorithm is a logical flow of thought that can be poured into the form of writing [4].

2.3 Boyer Moore

The Boyer-Moore algorithm performs comparisons starting from right to left, but the window shift remains from left to right [5]. The Boyer-Moore algorithm is one of the proper string searches and well-known matching algorithms used in single pattern matching and assumes very quickly in its performance. Boyer Moore's algorithm has been proven as one of the most efficient algorithms in string search applications using natural language (not binary language). This algorithm has often been implemented for the "Search" and "Substitute" functions of the text editor.

This algorithm is similar to Knuth Morris Pratt (KMP) algorithm where both algorithms will perform checking leaps in the string search process. But unlike the Knuth Morris Pratt algorithm, Boyer Moore's algorithm does a comparison pattern from right to left.

3 RESULT AND DISCUSSION

In typing is often the problem. One of them is often mistakes in terms of typing words that form a sentence. So to know which word is wrong in a text will require a little time or must read the word carefully to correct the wrong word and then will be justified manually. A text file is a file containing information in the form of text or data derived from a word processing application which will be used as a source of words to be checked or corrected. In the process of searching the wrong word in a text file using word repair application there are many possibilities that can cause the word can not be fixed, for example in a word inserted symbol or punctuation that is not appropriate in a word and the wrong word has not been stored On the data storage. Word search is a process of comparison between pattern and text. Where the pattern in word repair application is a text file that the text file content will be matched with the text. The process of searching words in a text file starts with the examination of the contents of the text file that has been first entered into the text box word repair application.

The process of examining the contents of a text file in finding the wrong word is done by comparing the entire contents of a text file with text already stored in the database, where the contents of the text file will be separated based on spaces so as to form a word of speech. The word that has been separated it will be matched with the existing word in the database. If there is a similarity between words searched with a matching word then the word is categorized into the wrong word then automatically the wrong word will be replaced with the correct word. To make it easier to find the wrong word, string matching technique is required. String matching or often called string matching is a technique for searching all occurrences of a short string pattern [0..m-1] called pattern in a long string of text [0..n -1] called text. One of the known string matching algorithms is Boyer Moore, so the word repair application that will be designed will use the Boyer Moore algorithm.



The Boyer Moore algorithm was created by R.M Boyer and J.S Moore. This algorithm is well known for being applied to many matching algorithms for multiple strings (multi pattern).

For example, there is a matching effort that occurs in the text $[i \dots j + n - 1]$, and assume the first mismatch occurs between text $[I + j]$ and pattern $[j]$, $0 < j < n$. Means, text $[i + j + 1..i + n - 1] = \text{pattern}[j + 1..n]$ and $a = \text{text}[I + j]$ is not the same as $b = \text{pattern}[j]$. If u is a suffix of a pattern before b and v is a prefix of the pattern, then possible shifts are:

1. The good-suffix shift consists of aligning the text pieces $[I + j + n - 1]$ pattern $[j + 1..n - 1]$ with the rightmost appearance on the pattern preceded by a character different from the pattern $[j]$. If there are no such chunks then the algorithm will align the suffix v from the text $[i..j + 1..i + n - 1]$ with the prefix of the same pattern.
2. A bad-character shift consisting of the alignment $[I + j]$ with the rightmost occurrence of the character in the pattern. If the character is not in the pattern, then the pattern will be aligned with the text $[i + n + 1]$.

Systematically, the steps performed by Boyer Moore's algorithm when matching strings are:

1. Boyer Moore algorithm, start matching the character at the beginning of the text.
2. From left to right, this algorithm will match characters per character in the corresponding text, until one of the following conditions is met.
 - a. In pattern and in comparable text match (mismatch).
 - b. All characters in a suitable pattern. The algorithm will notify the invention in this position.
 The algorithm then shifts the pattern by maximizing the good - suffix and bad - character shift values, then repeating step 2 until the pattern is in the text. An example of using Boyer Moore's algorithm for searching patterns in text:

Text = PERBAIKAN
 Pattern = PERBAKAN

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern	P	E	R	B	A	K	A	N						

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern		P	E	R	B	A	K	A	N					

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern			P	E	R	B	A	K	A	N				

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern				P	E	R	B	A	K	A	N			

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A



Pattern				P	E	R	B	A	K	A	N				
---------	--	--	--	---	---	---	---	---	---	---	---	--	--	--	--

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern					P	E	R	B	A	K	A	N		

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern						P	E	R	B	A	K	A	N	

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text	P	E	R	B	A	I	K	A	N		S	A	Y	A
Pattern							P	E	R	B	A	K	A	N

Make a pattern shift by paying attention to the matching text until he finds the same match, if not found then do shift as much as the character on the pattern.

Use case diagrams are graphical images of some or all of the actors. Use case diagrams explain the benefits of a system when viewed in the view of people who are outside the system. The diagram shows the functionality of a system or class how a system interacts with the outside world. Use case diagrams explain and explain the needs (requirement) desired or desired user (user), and very useful in determining the organizational structure and mode I of a system. Use case diagram describes an interaction between one or more actors with the information system to be created. Use case diagram is used to find out what functions are in a system. Here's a picture use case diagram in the text-fixing app:

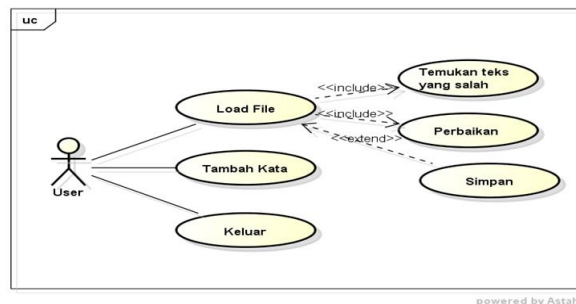


Figure 1: Use case diagram text improvement application

4 CONCLUSION

From the previous analysis, conclusions can be drawn, where the conclusions may be useful to the reader so that it can be more useful. The conclusions are as follows:

1. Improved text-enhanced apps can perform incorrect word searches on text files.
2. The Boyer Moore algorithm can be applied to text-based application design to perform incorrect search or string matching so that it can make it easier for users to find the wrong word and fix it automatically
3. Text application has been designed using Microsoft Visual Studio 2008 and MySQL as database.



REFERENCES

- [1] A. A. Alsaheel, A. H. Alqahtani, and A. M. Alabdulatif, "Analysis of Parallel Boyer-Moore String Search Algorithm," *Glob. J. Comput. Sci. Technol. Hardw. Comput.*, vol. 13, no. 1, pp. 43–47, 2013.
- [2] Y. Djahir and D. Pratita, *Bahan Ajar Sistem Informasi Manajemen.pdf*. Yogyakarta: Deepublish, 2015.
- [3] E. Utami and Sukrisno, *10 Langkah Belajar Logika dan Algoritma Menggunakan Bahasa C dan C++ di GNU/Linux*. Yogyakarta: Andi, 2005.
- [4] H. Pratiwi *et al.*, "Implementasi Algoritma Brute Force dalam Aplikasi Kamus Istilah Kesehatan," *J. Ilm. Teknol. Inf. Terap.*, vol. II, no. 2, pp. 119–125, 2016.
- [5] Kencana Wulan Argakusumah, "Implementasi Algoritma Boyer-Moore pada Aplikasi Kamus Kedokteran Berbasis Android," *Ultimatics*, vol. VI, no. 2, pp. 70–78, 2014.
- [6] A. Hadi, *Pemahaman dan Penerapan ISO/IEC 17025:2005 Persyaratan Umum Kompetensi, Laboratorium Pengujian dan Laboratorium Kalibrasi*. Jakarta: PT. Gramedia Pustaka Utama, 2007.
- [7] Y. A. dan R. F. Riyanarto Sarno, *Semantic Search*. Yogyakarta: Andi, 2014
- [8] F. T. Waruwu and M. Mesran, "IMPLEMENTASI ALGORITMA KNUTH MORRIS PRATT PADA APLIKASI KAMUS ISTILAH LATIN FLORA DAN FAUNA BERBASIS ANDROID," *Maj. Ilm. INTI*, vol. 4, no. 1, pp. 96–102, 2014.
- [9] F. T. Waruwu "Perbandingan Algoritma Knuth Morris Pratt Dan Boyer Moore Dalam Pencocokan String Pada Aplikasi Kamus Bahasa Nias" *J. Ilm. INFOTEK*, vol. 1, no. 1, 2016.
- [10] E. Rismawati, E. Buulolo, M. Sianturi, and F. T. Waruwu, "IMPLEMENTASI ALGORITMA KNUTH-MORRIS-PRATT DALAM PENCARIAN NOMOR DAN LIRIK LAGU DALAM KIDUNG JEMAAT BERBAHASA NIAS," *J. Ris. Komput.*, vol. 3, no. 4, pp. 10–15, 2016.