Knuth Morris Pratt - Boyer Moore Hybrid Algorithm for Knowledge Management System Model on Competence Employee in Petrochemical Company

1st Muhammad Riki Apriyadi
University of Sriwijaya
Palembang, Indonesia
m_riki@pusri.co.id

2nd Ermatita Ermatita
University of Sriwijaya
Palembang, Indonesia
ermatita@unsri.ac.id

4th Reza Firsandaya Malik
University of Sriwijaya
Palembang, Indonesia
rezafm@unsri.ac.id

3rd Dian Palupi Rini
University of Sriwijaya
Palembang, Indonesia
dprini@unsri.ac.id

***Abstract*—Knowledge Management System (KMS) is a system that allows a company to create, obtain, store and use knowledge optimally which is useful for improving company progress. Knowledge storage is useful for maintaining knowledge, therefore a string search method is needed to access the existing knowledge in KMS, so that if the knowledge is needed then other employees can easily get it. In this research a knowledge search method will be conducted in the form of employee competencies in KMS, to get it back a string search must be performed to make it easily accessible to other employees. The search method in this study is a combination of the search Algorithm Knuth Morris Pratt and Boyer Moore. In this study, the combined Knuth Morris Pratt-Boyer Moore (KMP-BM) Hybrid Algorithm is used for modeling the knowledge management system in determining employee competencies in petrochemical industry companies. Employee competency data are the result of competency assessment of the company which is then grouped by human resource development in KMS. The results of this study are looking for a method of string matching with the application of the KMP-BM Hybrid Algorithm on the KMS model to store various knowledge about employee competencies in the company. By using a combination of the existing advantages KMP and BM Algorithm, searching for KMS can be done more quickly and accurately.**

Keywords—KMS, Employee Competencies, Petrochemical Industry Companies, Knuth Morris Pratt-Boyer Moore (KMP-BM) Hybrid Algorithm.

I. INTRODUCTION

Along with the increasing era of knowledge, a successful company will need existing knowledge in companies that require competence which is also an important asset for the progress of the company. Employee competence is a requirement that is owned by employees to carry out a particular job in accordance with their fields and responsibilities. The competency is obtained through seminars, training, and certification. Then the development of human resources, assesses each employee so that employee placement can be determined based on the competencies of the employee. The employee's refresher process of promotion, rotation, transfer, appreciation, or resignation from the employee's competence cannot be completed properly then it will disappear. Thus, in this study, seeking knowledge about the competencies needed in the knowledge management system (KMS). To access knowledge in a knowledge management system a model is needed that can help speed up the search for knowledge. Therefore, this research will be conducted using Knuth Morris Pratt-Boyer Moore Hybrid Algorithm (KMP-BM Algorithm) modeling. This Algorithm is used to model knowledge search in a knowledge management system with matching strings that contain the advantages of Knuth's Morris Pratt Algorithm and Boyer Moore's Algorithm. Research in conducting string matching has been carried out by several researchers, among others [1] has conducted string matching research with the Knuth Morriss Pratt Fuzzy Algorithm on the Knowledge Management System model that stores information and research on content solutions and overcome them. Another study conducted by [2] conducted the Knuth Morris Pratt Algorithm in a mandolin-Indonesian translator application so that the process of searching translated words was easier. In research [3] Boyer Moore and Knuth Morris Pratt's comparative analysis in Book Title Search with an exponential comparison method that is calculated from the total amount of memory used and the time taken from each matching process, then produce Boyer Moore's research is the fastest in doing a search. Research by [4] has developed the Knuth Morris Pratt and Boyer Moore Algorithm for string searching so that it results in a better string Algorithm search. From the results of previous studies in this study will model the Knowledge Management System looking for strings in finding knowledge about employee competencies in petrochemical companies using the KMP-BM Hybrid Algorithm Model.

II. LITERATUREREVIEW

A. Knowledge Management System

Knowledge Management System is an operational activity in managing knowledge as an asset, where in various strategies there is the distribution of the right knowledge to the right people and in a fast time so that they can interact with each other, share knowledge and apply in their respective fields for the advancement of company performance [5]. So that all employees are expected to participate in improving personal quality through the implementation of the Knowledge Management System by

creating, storing, tracking, transferring and applying existing knowledge.

B. Employee Competencies in Petrochemical Company

Competence is the set of demonstrable characteristics and skills that enable, and improve the efficiency of, performance of a job [11].

Some scholars see competence as a combination of practical and theoretical knowledge, cognitive skills, behavior and values used to improve performance; or as the state or quality of being adequately or well qualified, having the ability to perform specific role. For instance, management competency might include systems thinking and emotional intelligence, and skills in influence and negotiation.

Every employee in a company is required to master competence in accordance with the placement in their respective fields in a company. This competency is also called Functional competencies Functional competencies, Functional competencies are job-specific competencies that drive proven high-performance, quality results for a given position. They are often technical or operational in nature (e.g., "backing up a database" is a functional competency) [12].

In companies engaged in the petrochemical industry that require some competence in functional fields that must be mastered [13], Water Treatment Plant, Demin Plant, Cooling Tower Plant, Gas Turbine Generator, Boiler Package Plant, Waste Heat Tolerance Plant, and Gas Measurement System.

C. Method of Search String Matching

Based on research that has been done [6], strings are arrays of characters (numbers, alphabet, or other characters) that are usually presented as arrays of data structures. Strings can be words, phrases, or sentences. Whereas a string that matches [6] is defined as a problem for finding character string pattern settings in strings or other parts of text content.

III. RESULT AND DISCUSSION

A. Research Methdology

This research will produce a KMS that uses Knuth Morris Pratt-Boyer Moore Hybrid Algorithm for Knowledge Management System Model on Competence Employee in Petrochemical Company, while the research methodology produced is as follows:

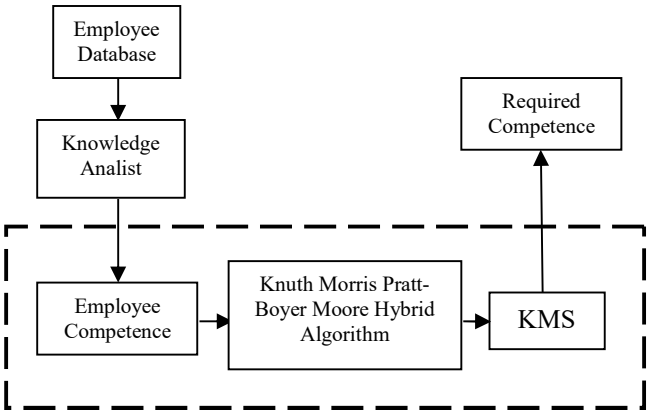


Fig. 1. Research Methodology

B. Knowledge Analist

This research will be carried out the mixing of the experience, values, contextual information and vision experts who provide an evaluation frameworkrelated to new experiences andinformation.

The Modeling knowledge management system is done by:

a) Knowledge discovery system: the search for knowledge of data and information.

b) Knowledge capture system: the process of gaining knowledge of explicit or tacit knowledge

c) Knowledge sharing system: the process of distributing knowledge [7].

d) Modeling Knowledge Management System with String Matching and KMP-BM Hybrid Algorithm Model.

C. Search String Matching

String Matching is the search for a pattern in a text [8]. The working principle of string matching Algorithm is as follows:

a) Scan text with the help of a window that is the same size as the pattern length.

b) Place the window at the beginning of the text.

c) Comparing the characters in the window with the characters of the pattern. After matching (whether the results match or not match), do the right shift on the window. This procedure is repeated until the window is at the end of the text. This mechanism is called a sliding - window.

String Matching is divided into two, namely exact string matching and heuristic (statistical matching). Exact matching is used to find patterns originating from one text. The exact matching Algorithm is classified into three parts according to the search direction, namely:

1. Search direction from left to right

The Algorithms included in this category are Brute Force, Morris and Pratt (which was later developed by Knuth, Morris and Pratt).

2. Direction of reading from right to left

The Algorithm included in this category is Boyer Moore which was later developed into the Turbo Boyer Moore Algorithm, Tuned Boyer Moore and Zhu Takaoka.

3. The search direction is determined by the program

The Algorithms included in this category are the Colussi and Crochemore-Perrin Algorithms.

Heuristic matching is a technique used to connect two separate data when exact matching is unable to overcome because of restrictions on available data. Heuristic matching can be done by calculating the distance between the pattern and the text. Exact and heuristic matching has similar meanings, but differ in writing. Other texts and patterns. Text (text) is a (long) string whose length is n. The pattern is a

string of length m characters (m <n) that will be searched for in text [9].

D. Knuth-Morris-Pratt Algorithm

Knuth-Morris-Pratt Algorithm was developed by D. E. Knuth, together with J. H. Morris and V. R. Pratt [3]. The Knuth-Morris-Pratt Algorithm is a development of the previous string search Algorithm, the Brute Force Algorithm. KMP (Knuth Morris Pratt) Algorithm is an Algorithm used to do string matching. This Algorithm is a type of Exact String Matching Algorithm which is a matching string exactly with the arrangement of characters in a matched string that has the number or sequence of characters in the same string. There are several definitions in the Algorithm KMP:

1. Let A be the alphabet and x = x1x2 ... xk is a string whose length k is formed from the characters in the alphabet A.
2. The prefix of x is the u string (substring) u with u = x1x2 ... xk - 1, k ∈ {1, 2, ..., k - 1} in other words, x starts with u.
3. The suffix of x is the u string (substring) u with u = xk - b xk - b + 1...xk, k ∈ {1, 2,..., k - 1} in other words, x ends with v.
4. The border of x is the string r such that r = x1x2 ... xk - 1 and u = xk - b xk - b +1 ... xk, k ∈ {1, 2, ..., k − 1}, in other words, the edges of x are string-forms which are both the prefix and also the actual ending of x.

Periphery function b (j) is defined as the longest prefix measure of P which is the suffix of P [1..j].

The Knuth Morris Pratt Algorithm is not like the Brute Force Algorithm that matches strings by checking and shifting every single character, in the Knuth Morris Pratt Algorithm the pattern incompatibility information with the text is stored to determine the number of shifts. So that the Knuth Morris Pratt Algorithm makes further shifts according to the information stored, which causes the search time to be significantly reduced, here are the work steps of the Knuth Morris Pratt Algorithm:

1. The Knuth Morris Pratt's Algorithm starts matching patterns at the beginning of the text.
2. From left to right, this Algorithm will match characters per character pattern with the characters in the corresponding text, until one of the following conditions is met:
   a. Characters in the pattern and in the text being compared do not match (mismatch).
   b. All characters in the pattern match. Then the Algorithm will notify the discovery in this position.
   c. The Algorithm then shifts the pattern based on the next table, then repeats step 2 until the pattern is at the end of the text. Next is the process of applying the Algorithm

Knuth Morris Pratt to look for strings in a string, while the process looks as follows:

1. Known S string variables with an array of letters as follows:

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R

2. Known pattern P is the word to be searched for in the variable S:

H	E	A	T

3. The first step is to compare the pattern P1 with the string S2, here are the results;

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
H	E	A	T													

Pattern 1 does not match string 1 then the pattern will shift one position to the right.

4. The second step compares pattern P1 with string S2, here are the results.

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
	H	E	A	T												

Pattern 1 does not match string 2 then pattern will shift one position to the right.

5. The third step compares pattern P1 with string S3, here are the results

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
		H	E	A	T											

Pattern 1 does not match string 3 then the pattern will shift one position to the right.

6. The fourth compare pattern P1 with string S4, here are the results

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
			H	E	A	T										

Pattern 1 does not match string 4 then the pattern will shift one position to the right.

7. The fifth step compared pattern P1 with string S5, here are the results.

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
				H	E	A	T									

Pattern 1 does not match the string5then the pattern will shift one position to the right.

8. The sixth step compared pattern P1 with string S6, here are the results.

W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
					H	E	A	T								

Pattern 1 does not match the string6 then the pattern will shift one position to the right.

9. The Seventh step compares the pattern P1 with string S7, here are the results.

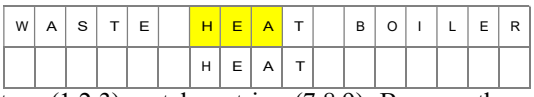
W	A	S	T	E		H	E	A	T		B	O	I	L	E	R
						H	E	A	T							

Pattern 1 matches string 7. Because there is a match, the Knuth Morris Pratt Algorithm will save this information, and the pattern will not shift and continue matching pattern 2 with string 8.

10. The Eighth stepcompares pattern P (1,2) with string S (7,8), here are the results.

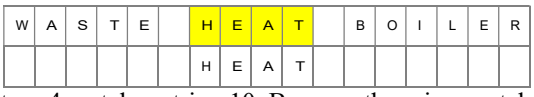
Pattern (1,2) matches string (7,8). Because there is a match, the Knuth Morris Pratt Algorithm will store this information, and the pattern will not shift and continue matching pattern (3) with string (9).

11. The Ninth stepcompares pattern P (1,2,3) with string S (7,8,9), here are the results.



Pattern (1,2,3) matches string (7,8,9). Because there is a match, the Knuth Morris Pratt Algorithm will store this information, and the pattern will not shift and continue matching pattern 4 with string 10.

12. The Tenth stepcompared pattern P (1,2,3,4) with string S (7,8,9,10), here are the results



Pattern 4 matches string 10. Because there is a match, the Knuth Morris Pratt Algorithm will store this information, and the pattern will not shift and continue matching pattern 5 with string 11. However, because the number of patterns is only 4 letters, the search will be stopped and the result is that the pattern P matches the S string by 100 percent.

E. Boyer Moore Algorithm

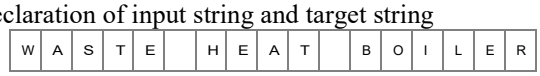
Boyer Moore's Algorithm was published by Robert S. Boyer, and J. Strother Moore in 1977. Matching characters using the Boyer Moore Algorithm start of the string to the right of the pattern. The idea behind this Algorithm is to start matching characters from the right, so more information will be obtained [10].

The steps to find the string in Boyer Moore's Algorithm are as follows:

1. First, we need 2 tables with the Match Heuristic (MH) and Occurrence Heuristic (OH) approach to determine the number of shifts that will be made in the pattern (P) if a character is found that does not match the matching process of the characters in the text (S).
2. If the comparison process, there is a mismatch of characters between the characters in P and S, then the shift is done by looking at the two tables with the greatest shift value selected.
3. The possible settlement in shifting P is that if there is no matching character in the previous match, the shift is done by looking at the value of the shift in the occurrence heuristic table. If the character being compared does not exist in P then the shift is performed as many as the number of characters contained in P, but if the mismatched character is contained in the P string, then the shift is carried out based on the table.
4. If the character in the text being compared matches the character in the string.

Next is the process of applying Boyer Moore's Algorithm to find the target string from the input string.

1. Declaration of input string and target string



Then the OH and MH tables are determined as below:

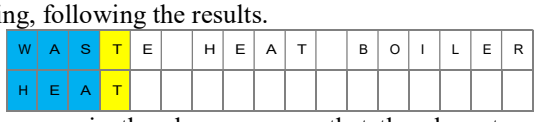
Table I. *Occurrence Heuristic* (OH) Table

Index	0	1	2	3
Target	H	E	A	T
OH	3	2	1	0

Table II. *Match Heuristic* (MH) Table

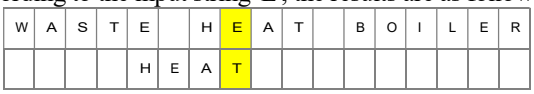
Index	0	1	2	3
Target	H	E	A	T
MH	4	4	4	1

2. First step, the last character of the target string ‘T’ matches the ‘T’ character of the word ‘WASTE HEAT BOILER’, then tests all the characters contained in the target string as many as n target strings from the input string, following the results.



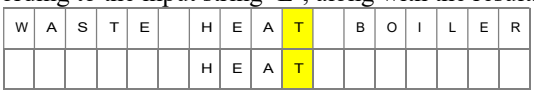
It appears in the above process that the characters that match are only the T characters and not for other characters, then the next process of the target string advances as many as n the target string characters.

3. Second step, test the last character of the target string 'T' according to the input string 'E', the results are as follows.

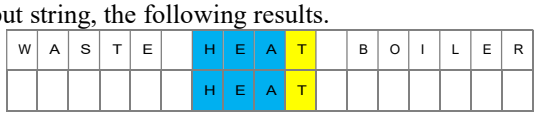


It appears in the above process, that the target string character 'T' does not match the character string 'E', then the next process is to compare the largest OH and MH values, obtained by shifting by 2 characters.

4. Third step, test the last character of the target string ‘T’ according to the input string ‘E’, along with the results.



In the above process, it appears that the last character of the target string ‘T’ is the same as the input string ‘T’ for the same position, after that all the same character positions are tested between the target string and the input string, the following results.

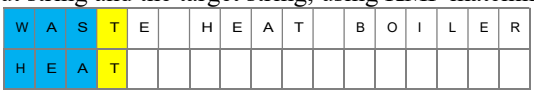


Appear in the picture is appropriate, then the process is stopped and the word "HEAT" is found in the fourth step up to the sixth step in characters 7 through 10, then the remaining characters from the input string are not processed.

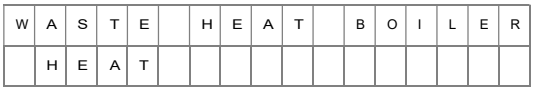
F. Knuth Morris Pratt-Boyer Moore Hybrid Algorithm

Following is the process of applying the Knuth Morris Pratt-Boyer Moore Hybrid Algorithm [4], note that the input string is 'WASTE HEAT BOILER' and the target string is 'HEAT'.

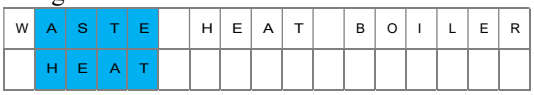
1. The first step is to check the exact position between the input string and the target string, using KMP matching.



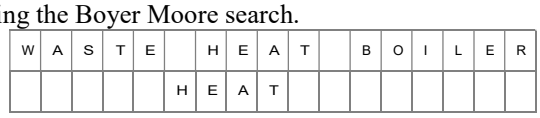
Because there is only one character in common between the target string and the input string, 'T', the input string is shifted 1 position to the right and the results are as follows.



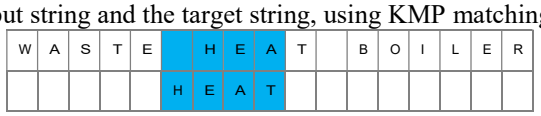
2. The second step is to check the exact position between the input string and the target string, using KMP matching.



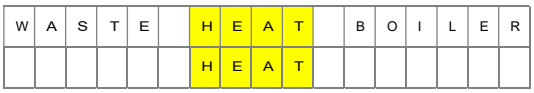
On the shift there are no similar characters, the target string shifts as much as the length of the target string by using the Boyer Moore search.



3. The third step is to check the exact position between the input string and the target string, using KMP matching.



Seen in the above process, there are some of the same characters between the target string and the input string, based on the Boyer Moore Algorithm function, if there are the same characters, the position of the characters between the target string and the destination string must be matched as below.



In this step, for each character in the target string according to the input string, the process is stopped because the string has been found. Note the difference in string matching done using Boyer Moore and KMP Algorithms, the process is faster because it only matches between characters and not for all the characters in the target string.

IV. CONCLUSION

Knuth Morris Pratt-Boyer Moore Hybrid Algorithm is an Algorithm that combines Knuth-Morris-Pratt and Boyer Moore Algorithm. Combining the strengths of Knuth Morris Pratt Algorithm and Boyer Moore Algorithm will result in a faster and more efficient search process for employee competencies in petrochemical companies. So that the results of this modeling can be the basis for developing a Knowledge Management System.

ACKNOWLEDGMENT

The Authors thank to Universitas of Sriwijaya and PT.Pupuk Sriwidjaja for facilitateing research data.

REFERENCES

[1] E. Ermatita and D. Budianta, "Fuzzy Knuth Moris PrattAlgorithm for knowledge management system model on knowledge heavy metal content in oil plants," *ICECOS 2017 - Proceeding 2017 Int. Conf. Electr. Eng. Comput. Sci. Sustain. Cult. Herit. Towar. Smart Environ. Better Futur.*, pp. 188–192, 2017.

[2] P. B. Mandailing-indonesia, R. K. Hondro, Z. A. Hsb, and R. D. Sianturi, "APLIKASI PENERJEMAHAN BAHASA MANDAILING-INDONESIA," no. January, 2016.

[3] A. Fau, Mesran, and G. L. Ginting, "Analisa Perbandingan Boyer Moore Dan Knuth Morris Pratt Dalam Pencarian Judul Buku Menerapkan Metode Perbandingan Eksponensial (Studi Kasus : Perpustakaan STMIK Budi Darma)," *J. Times (Technology Informatics Comput. Syst.*, vol. 6, no. 1, pp. 12–22, 2017.

[4] R. Y. Tsarev, A. S. Chernigovskiy, E. A. Tsareva, V. V. Brezitskaya, A. Y. Nikiforov, and N. A. Smirnov, "Combined string searching Algorithm based on knuth-morris- pratt and boyer-moore Algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 122, no. 1, 2016.

[5] A. M. Abubakar, H. Elrehail, M. A. Alatailat, and A. Elçi, "Knowledge management, decision-making style and organizational performance," *J. Innov. Knowl.*, vol. 4, no. 2, pp. 104–114, 2019.

[6] M.SyaroniandR.Munir,"PencocokanStringBerdasarkanKemiripan Ucapan (Phonetic String Matching) dalam Bahasa Inggris," vol. 2, pp.1–6.

[7] I. Becerra-Fernandez and J. SABHERWAL, "Knowledge Management: Systems and Processes, New York: ME Sharpe." Inc, 2010.

[8] P. P. Dinamis, "Algoritma Fuzzy String Matching untuk Pencocokan String," 2019.

[9] M. A. S. Naser, N. A. Rashid, and M. F. Aboalmaaly, "Quick-Skip Search Hybrid Algorithm for the Exact String Matching Problem," *Int. J. Comput. Theory Eng.*, vol. 4, no. 2, pp. 259–265, 2012.

[10] G. L. Ginting, "Implementasi Algoritma Boyer-Moore Pada Aplikasi Pengajuan Judul Skripsi Berbasis Web," *Pelita Inform.*, 2014.

[11] Wikipedia, http://id.wikipedia.org.

[12] M. Y. A. Alsabbah and H. I. Ibrahim, "Employee Competence (Soft and Hard) Outcome of Recruitment and Selection Process," vol. 3, pp. 67–73, 2013.

[13] Marianti, "Plant Utiltas PT. Pupuk Sriwidjaja Palembang," 2018.

978-1-7281-2930-3/19/$31.00 ©2019 IEEE