Application of knuth-morris-pratt algorithm on web based document search

Amirul Fatah1, Isturom Arif1, Fauzi Farchan2, Vieronica Varbi Sununianti2, Riski Amalia Madi3, Erwingsyah Satria4, Endang Fourianalistyawati5, Irwan Bempah6, Dwi Ermayanti Susilo7, Early Ridho Kismawadi8, Nopriadi9, Ruzita Sumiati10, Iif Novita Sari11, Fajar Kusnadi Kusumah Putra12, Hasina Fajrin13, Ebin E Danius14, Purwo Subekti15, Yuska Noviyanty16, Nurdianna Siregar17, Nur Fauziah Siregar18, Rudi Abdullah19, Hardianto Djanggih20, Sri Puspa Dewi

1Department of Informatics Engineering, Universitas Maarif Hasyim Latif, Indonesia
2Universitas Wiralodra Indramayu, Indonesia
3Sriwijaya University, Indonesia
4Universitas Halu Oleo, Indonesia
5Universitas Bung Hatta, Indonesia
6Universitas YARSI, Indonesia
7Universitas Negeri Gorontalo, Indonesia
8STIE PGRI Dewantara Jombang, Indonesia
9Institute for Islamic Studies Langsa IAIN Langsa, Indonesia
10Universitas Putera Batam, Indonesia
11Politeknik Negeri Padang, Indonesia
12Universitas Kanjuruhan Malang, Indonesia
13Hotel Administration Department, STP NHI Bandung, Indonesia
14Balai Bahasa Sulawesi Selatan, Indonesia
15Universitas Hein Namotemo, Indonesia
16Universitas Pasir Pengaraian, Riau, Indonesia
17Akademi Farmasi Yayasan alfatah, Indonesia
18Universitas Nahdlatul Ulama Sumatera Utara, Indonesia
19IAIN Padangsidimpuan Indonesia
20Universitas Muhammadiyah Buton, Indonesia
21Universitas Tompotika Luwuk Banggai, Indonesia

amirul.fatah@gmail.com

Abstract. The rapid development of information technology makes a number of documents in a university need to be kept in web-based storage to be utilized by the entire academic community. One example is a thesis document of Universitas Maarif Hasyim Latif students who do not have a web-based storage system so that the academic community has difficulty in finding lecture references at the university. System management is needed to be able to handle student thesis document reference searches. In searching for lecture references, a document search algorithm is needed, the algorithm that can be applied to the system made is Knuth-Morris-Pratt and the development of the system using Waterfall. The results of system development show that the system that has been made successfully knows the position of the words entered by the user so
that it is very useful for the entire academic community in searching for thesis documents for students, especially students who are currently undergoing a thesis, has been guided by a supervisor so that the University can find out the ratio of a supervisor in conducting guidance to students.

1. Introduction

So far, the softcopy documents on CD aren't widely employed by the educational community as a result of they're solely held onto into stacks by the chair of the study program and are liable to physical injury from CDs in order that necessary documents from the students' theses aren't used. Thus the requirement for storage and rummage around for web-based student thesis documents in order that it are often employed by the educational community, particularly students United Nations agency are taking theses and document management is straightforward. The absence of a web-based student thesis document storage hampers the educational community from sorting out thesis document references, particularly students United Nations agency take theses. In looking out documents, what's required could be a String Matching / matching string and Pattern / pattern within the student's thesis document [1].

String matching is used to find a string called Pattern in a string called text [2]–[4]. String matching or often referred to as string matching is the process of finding all occurrences of queries, which is then called a pattern into a longer string (text). The pattern is denoted by \( x = x[0..m-1] \) and the length is \( m \). The text is denoted by \( y = y[0..n-1] \) and the length is \( n \). Both strings consist of a set of characters called alphabets that are represented by \( \Sigma \) and have a size. String matching is divided into two, namely exact matching and heuristic [5]–[8]. One example of a String matching algorithm is Knuth-Morris-Pratt. Knuth-Morris-Pratt is a string matching process. If there is a discrepancy when the pattern is parallel to the text \( [i..i+n-1] \), we can assume the first discrepancy occurs between the text \( [i+j] \) and pattern \( [j] \), with \( 0 < j < n \). Means, the text \( [i..i+j-1] = \) pattern \( [0..j-1] \) and a = text \( [i+j] \) are not the same as b = pattern \( [j] \) [8]–[10]. Ramadhani (2017), has already applied Knuth-Morris-Pratt Algorithm to Book Search Application Design in Darul Ilmi Murni Islamic International School Library with Knuth-Morris-Pratt Algorithm which results in searching for book data in libraries to be easier, faster and also efficient. Research on the Knuth-Morris-Pratt algorithm was carried out by Usman Ependi and Nia Oktaviani with the Black Box (2017) testing approach, namely "Abstract Keyword Searching with Knuth Morris Pratt Algorithm" that the device build has succeeded in finding 100% the same as the keywords that have been entered by users in searching for articles. Jialamm Igaph Sinaga, Mesran, Efori Buulolo also conducted research on "Mobile word search application on the meaning of the Android-based verse of the Qur'an using string matching algorithm" which results in the knuth morris pratt algorithm can be applied in the design of a word search mobile application in the meaning of the verse Al-Qur'an so that it can make it easier for users to search for the word you want to translate [11].

2. Methods

The research was conducted at Universitas Ma'arif Hasyim Latif which was located at Jalan Raya Ngelom Megare No. 30 Taman Sidoarjo and the time of the study began on March 19, 2018 to July 15, 2018. The object of the study was the process of searching for student thesis documents consisting of titles, abstracts, authors with the Knuth-Morris-Pratt algorithm. The involved actors consist of admin, lecturers, staff and students.

3. Result and Discussions

To ensure the system is made with the correct function, it needs to be tested using the Black Box test as in the table below:

| No. | Components                               | Test type     |
|-----|------------------------------------------|---------------|
| 1   | Searching for thesis document            | Black Box     |

Table 1. Forms of testing
In the student thesis document search function, to test it the user must enter the word / pattern in the search column then by clicking the search button to the right of the search column. After that, the system will process the search using the Knuth-Morris-Pratt algorithm. If the word entered has been found, the system will inform its position according to the word that has been entered as shown below.

Figure 1. Testing the search function with the Knuth-Morris-Pratt algorithm.

From the test results obtained, information about the position of words and documents containing the words that have been entered so that the system has successfully implemented the Knuth-Morris-Pratt algorithm in searching for student thesis documents and knowing the position of words in the database. Knuth-Morris-Pratt algorithm testing for searching patterns in text:

Text = Tenaga, Menikah
pattern = Menikah
peripheral / leap pattern = Menikah

Table 2. Fringe Pattern Function

| Pattern | n | i | K | a | h |
|---------|---|---|---|---|---|
| Peripheral | 0 | 0 | 0 | 0 | 0 |

Solution:

Table 3. Matching Knuth-Morris-Pratt to 1st

| text | t | e | n | A | g | a | , | m | e | n | i | k | a | h |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pattern | n | i | k | A | h |
| index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |

Description: pattern 1 does not match text 0, then one step is shifted right to the next index.

Table 4. Matching Knuth-Morris-Pratt to 2nd

| text | t | e | n | A | g | a | , | m | e | n | i | k | a | h |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pattern | n | i | K | a | h |
| index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |

Description: pattern 1 does not match text 1, then one step is shifted right to the next index.
Table 5. Matching Knuth-Morris-Pratt to 3rd

| Pattern  | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Text     | t  | e  | n  | A  | g  | a  | .  | m  | e  | n  | i  | k  | a  | h  |    |

Description: pattern 1 matches text 2, but pattern 2 does not match text 3, then shift 2 steps to the right to the next index.

The matching method of Knuth-Morris-Pratt is run up to the ordinal loop, with the results of pattern one matching text ten, the Knuth-Morris-Pratt rule can store this data and therefore the pattern won't shift, then check pattern two with text eleven apparently appropriate then the Knuth-Morris-Pratt rule can store this data and therefore the pattern won't shift, except check pattern three with text twelve apparently appropriate then the Knuth-Morris-Pratt rule can store this data and therefore the pattern won't shift, still check the pattern four with text thirteen apparently appropriate then the Knuth-Morris-Pratt rule can store this data and therefore the pattern won't shift, except check the pattern five with text fourteen seems to be appropriate then the Knuth-Morris-Pratt rule can store this data and therefore the pattern won't do shift, as a result of the text doesn't exist, the Knuth-Morris-Pratt rule can stop and therefore the position is tiny let all the patterns be told. in order that the pattern = wedding is found within the text = power, married at ten,11,12,13,14. However, if the search results don't exist, the system can inform the info not found within the info, as shown below.

Figure 2. Document search none.

4. Conclusions

The system created will offer convenience for users in managing student thesis documents that are manually keep. The system created is correct to work out the position of the word entered by the user. Word search within the info has succeeded in knowing the position of the word that has been entered in order that the Knuth-Morris-Pratt rule will facilitate to search out student thesis documents. The system created has succeeded in displaying info regarding the quantity of scholars UN agency area unit guided by a supervisor.

References

[1] R. Rahim et al., “Prototype File Transfer Protocol Application for LAN and Wi-Fi Communication,” Int. J. Eng. Technol., vol. 7, no. 2.13, pp. 345–347, 2018.
[2] U. Ependi and N. Oktaviani, “Abstract Keyword Searching with Knuth Morris Pratt Algorithm,” Sci. J. Informatics, vol. 4, no. 2, pp. 150–157, 2017.
[3] D. Shapira and A. Daptardar, “Adapting the Knuth–Morris–Pratt algorithm for pattern matching in Huffman encoded texts,” Inf. Process. Manag., vol. 42, no. 2, pp. 429–439, 2006.
[4] K. K. Mandumula, “Knuth-Morris-Pratt,” 2011.
[5] D. Ramadhani, “PERANCANGAN APLIKASI PENCARIAN BUKU PADA PERPUSTAKAAN ISLAMIC INTERNATIONAL SCHOOL DARUL ILMI MURNI DENGAN ALGORITMA KNUTH MORRIS PRATT,” Pelita Inform. Inf. dan Inform., vol. 16, no. 1, 2017.
[6] M. Régnier, “Knuth-Morris-Pratt algorithm: an analysis,” in International Symposium on
Mathematical Foundations of Computer Science, 1989, pp. 431–444.

[7] P. Gawrychowski, A. Jeż, and Ł. Jeż, “Validating the Knuth-Morris-Pratt failure function, fast and online,” Theory Comput. Syst., vol. 54, no. 2, pp. 337–372, 2014.

[8] R. Guha, R. McCool, and E. Miller, “Semantic search,” in Proceedings of the 12th international conference on World Wide Web, 2003, pp. 700–709.

[9] J. Zhong, H. Zhu, J. Li, and Y. Yu, “Conceptual graph matching for semantic search,” in International Conference on Conceptual Structures, 2002, pp. 92–106.

[10] D. E. Knuth, J. Morris James H, and V. R. Pratt, “Fast pattern matching in strings,” SIAM J. Comput., vol. 6, no. 2, pp. 323–350, 1977.

[11] J. I. Sinaga and E. B. Mesran, “APLIKASI MOBILE PENCARIAN KATA PADA ARTI AYAT AL-QUR’AN BERBASIS ANDROID MENGGUNAKAN ALGORITMA STRING MATCHING.”