Abstract Keyword Searching with Knuth Morris Pratt Algorithm

Usman Ependi[1], Nia Oktaviani[2]

[1,2]Faculty of Computer Science, Universitas Bina Darma, Indonesia
Jl. Jenderal Ahmad Yani No 3 Plaju Palembang, Indonesia
Email: u.ependi@binadarma.ac.id[1], niaoktaviani@binadarma.ac.id[2]

Abstract

This research was conducted to answer the problems of researchers for finding publication or papers that suitable with their topics. In this research was developed software for searching abstract keyword using Knuth Morris Pratt algorithm to answer the problems of researchers. Waterfall model used to develop abstract keywords searching software as tools development that has five phases, namely communication, planning, modeling, construction and deployment. The software was developed can display search results effectively and efficiently according to the enter search keywords, it can see while search results are shown is in case sensitive. The software is also tested; the testing process is conducted by functional observation with black box testing approach, Observation results while testing is conducted show the software running suitable with expected or 100% same with entered keyword, so worthy to be used as one of the tools of researchers searching articles.

Keywords: *Knuth Morris Pratt, Abstract, Searching*

1. INTRODUCTION

Publication or scientific writing is the work of a scientist (who is the result of development) who wants to develop the science, technology and art gained through bibliography, collection of experiences, and previous knowledge of others. Publication or articles: scientific attitude statement of researchers. The purpose of publication: so that the author's ideas can be learned, then supported or rejected by the reader. The function of scientific papers is as a means to develop science, technology, and art [1]. Currently at any higher education both universities, high schools and colleges of higher learning all do research to produce publications of either journals, proceedings or other research results. Each researcher read each other's publication results to be used as a reference in doing research. Publication has an abstract as a short article which contains a brief overview of research activities and activities. Abstracts are usually placed at the beginning of a publication or research report as a prefix information for the readers.

In the process of searching for the publication of a researcher, both students and lecturers often have trouble finding articles of publication that suitable with their needs. researcher searches the publication on journal systems as well as in search engines. The difficulty is usually due to the less accurate search results in displaying the results. It condition because of a lot of publication are displayed not in accordance with the keywords sought, So researchers waste a lot of time for searching relevant publication that match what they want. To facilitate the process of publication search it is necessary to apply effective and efficient search techniques. One search technique that is suitable

for string search is Knuth Morris Pratt (KMP) algorithm. Step of KMP algorithm works starts by matching the pattern at the beginning of the text. From left to right, this algorithm will match the characters per character pattern with characters in the corresponding text, until one of the conditions is met [2]. The use of KMP algorithm in the process of searching the keyword abstract publication will greatly help the researchers because the search process and search results are displayed in accordance with what is desired. Because of the KMP algorithm, keep of the information used to perform the number of shifts in the search process and use that information to make a further shift, not just one character [3].

To test and know whether KMP algorithm can search effectively and efficiently then in this research conducted implementation of KMP algorithm in process of searching keyword of abstract publication. The implementation process is conducted by translating the KMP algorithm into the software. The KMP algorithm is one of the most efficient algorithms in searching [4]. So the KMP algorithm is very compatible with any string search type [5]. Beside of that condition KMP Searching algorithms in computer science are used in many applications and their complexity analysis is important. More crucially, reducing the computation time heals the success of the application itself especially in the big data [6].

2. METHODS

2.1. Research Method

The research method used in this research is using descriptive research method. Descriptive research method is one of the many research methods used in research that aims to explain an event. Descriptive study is a study that aims to provide or describe a state or phenomenon that occurs today by using scientific procedures to answer the problem actually [7].

2.2. Development Method

In the process of implementing KMP algorithm in publication search then the development method used is waterfall, waterfall is a classical model that is systematic, sequential in building software. Phases in the waterfall model [8] as shown in Figure 1.
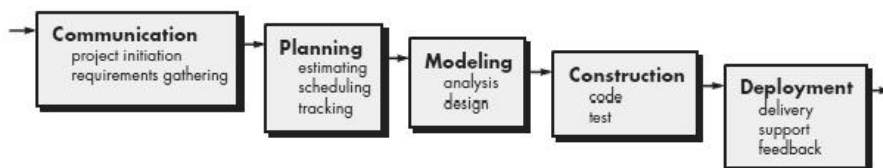

Figure 1. Waterfall Model [8]

a.   Communication, This step is an analysis of the needs of the software, and the stage to hold data collection by meeting with customers, or collect additional data either in journals, articles, or from the internet.

b. Planning, planning process is a continuation of communication process (analysis requirement). This stage will generate user requirement document or can be said as data related to user desire in making software, including scope, cost and time.

c. Modeling, the modeling process will translate the requirement to a predictable software design before coding. This process focuses on the design of data structures, software architectures, interface representations, and procedural (algorithmic) details.

b. Construction, Construction is the process of coding. Coding is a translation of design in a language that can be recognized by a computer. The programmer will translate the transaction requested by the user. This stage is a real step in working on software, meaning that the use of computers will be maximized in this stage. After the coding is complete it will be testing the system that has been made earlier. The purpose of testing is to find errors on the system for later repair.

c. Deployment, Deployment is the final stage in software development. After the analysis, design and coding process the ready-to-use system will be used by the user and subsequent periodic maintenance of the software is performed.

2.3 Testing Method and Data

Testing is one of the processes in software development. Testing aims to see how software works. To do the testing used black box testing as a test method. Black Box Testing is a testing technique without reference to the internal structure of the software. In Black Box Testing does not require an expert in the field of programming when performing testing [9]. In the process of testing software for searching abstract keyword using Knuth Morris Pratt data algorithm used sourced from existing research data on if.binadarma.ac.id/sipi can be seen in Figure 2. But the data can be sourced from different places such as the online journal system or other indexed journals.

Figure 2. Sample of Data Source

3. RESULT AND DISCUSSION

Search software is software that can be used specifically to search publication either journal, proceedings or other research result. This software was developed to answer the problems that occur in researchers in search of publications as a research reference. The software has the main features of publishing searches either journals, proceedings or other research based on titles, abstracts and abstract keywords. The software has been developed using waterfall model with five (5) development phase i.e. communication, planning, modeling, construction and deployment. Searching process performed on the software can be seen in Figure 3.

Figure 3. Searching Form

To perform a publication search according to the keywords entered as shown in Figure 3; the Knuth Morris pretty algorithm is used as a tool in the search process that has been translated into the programming code as shown in Figure 4.

```
<?php
class KMP{
  function KMPSearch($p,$t){
    $result = array();
    $pattern = str_split($p);
    $text    = str_split($t);

    $jump = $this->preKMP($pattern);
      $i = $j = 0;
    $num=0;
    while($j<count($text)){
      while($i>-1 && $pattern[$i]!=$text[$j]){
        $i = $jump[$i];
      }
      $i++;
      $j++;
      if($i>=count($pattern)){
        $result[$num++]=$j-count($pattern);
        $i = $jump[$i];
      }
    }
   return $result;
  }
  function preKMP($pattern){
    $i = 0;
    $j = $jump[0] = -1;
    while($i<count($pattern)){
      while($j>-1 && $pattern[$i]!=$pattern[$j]){
        $j = $jump[$j];
      }
      $i++;
      $j++;
      if($pattern[$i]==$pattern[$j]){
        $jump[$i]=$jump[$j];
      }else{
        $jump[$i]=$j;
      }
    }
    return $jump;
  }
}
?>
```

Figure 4. Translation of KMP Algorithm to Code

In the search process as shown in Figure 4 the Knuth Morris Pratt algorithm is called to perform the search process. The steps that the Knuth-Morris-Pratt algorithm performs when matching the process as follows:
a. The Knuth-Morris-Pratt algorithm begins to match the pattern at the beginning of the text. From left to right, this algorithm will match characters per character pattern with characters in the corresponding text, until one of the following conditions is met:
   i. The characters in the pattern and in the text that are mismatch.
   ii. All characters in the pattern match. Then the algorithm will notify the invention in this position.
b. The algorithm then shifts the pattern by string, then repeats step 2 until the pattern is at the end of the text.

The KMP algorithm finds all occurrences of a pattern of length n in the text of length *m* with the time complexity of $O\,(m+n)$. KMP algorithm requires only $O\,(n)$ space of internal memory if text is read from external file. All O quantities do not depend on the size of the alphabet space. To clarify the process in the KMP Algorithm when performing the search process can be seen in Figure 5.

```
j = 0;
for (i = 0; i < n; i++)
for (;;) {        // loop until break
    if (T[i] == P[j]) { // matches?
    j++;          // yes, move on to next state
    if (j == m) {  // maybe that was the last state
        found a match;
        j = overlap[j];
    }
    break;
    } else if (j == 0) break;   // no match in state j=0, give up
    else j = overlap[j];     // try shorter partial match
}
```

Figure 5. Pseudo-code KMP Algorithms [10]

The steps of Knuth Morris Pratt algorithm can be illustrated as shown in Figure 6 based on pseudocode in figure 5. In the simulation process it can be seen that the search is done starting from string to *0* up to string to *n*. Provided that a fixed search is performed if the mismatch condition is up to string *n*.

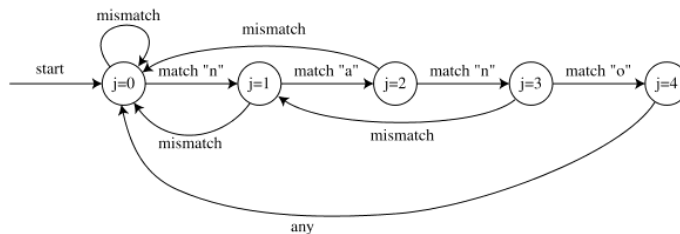**Figure 6.** Finite Automata KMP Algorithm [11]

After the search is done then the software will display search results based on keyword entered. The search results are case sensitive (100%) exactly same with keywords or the algorithm distinguishes between uppercase and lowercase, So the search results shown are exactly the same as the keywords as shown in Figure 7.
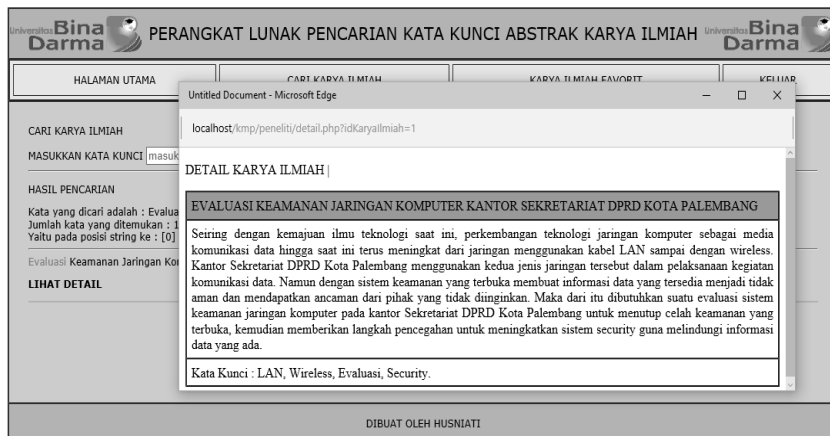

Figure 7. Search Result

To ensure that the software has been developed properly functionally it is necessary to test using black box testing with the test plan as in Table 1.

Table 1. Testing Plane

No	Component	Testing Type
1	Search Article	Black Box
2	Favorite Article	Black Box

From the test plan as in Table 1, the test results can be seen in Table 2

Table 2. Testing Result

No	Component	Expectation	Result
1	Search Article	Software can perform searches based on the keywords entered	Software is capable of displaying search results based on 100% keywords
2	Favorite Article	Software can make the favorite article of the search results based on user action	Software is able to make the favorite article search results based on user action

From the test results as in Table 2 can be said that the software using Knuth Morris Pratt algorithm running as expected (100% working as expected) and can serve as a tool for researchers in searching for the publication of both journals, proceedings and research reports.

4. CONCLUSION

The Knuth Morris Pratt algorithm can perform an abstract keyword search effectively and efficiently proven when the search process displays the results based on the keyword is case sensitive (100% same with entered keyword), so it can be used as a tool by researchers in finding publications such as journals, proceedings or research report. The software has also been developed with a waterfall model and performed black box testing with the test results show the functional software running well as expected.

5. REFERENCES

[1] Dwiloka, Bambang and Riana, Rati. 2005. *Teknis Menulis Karya Ilmiah*. Semarang: Rineka Cipta.

[2] Sunni, Ismail. 2010. Music-Finder Menggunakan Algoritma KMP Extension. *Makalah IF3051 Strategi Algoritma*.

[3] Wibowo, T., Wibowo, A. and Sari, R.P., 2012. Pembuatan Aplikasi untuk Mendeteksi Kebenaran Perintah Sql Query Menggunakan Metode Knuth-Morris Pratt (KMP). *Jurnal Aksara Komputer Terapan*, Vol 1(2).

[4] Lin, K.J., Huang, Y.H. and Lin, C.Y., 2013. Efficient parallel knuth-morris-pratt algorithm for multi-GPUs with CUDA. In Advances in Intelligent Systems and Applications Volume 2 (pp. 543-552).

[5] Tsarev, R.Y., Chernigovskiy, A.S., Tsareva, E.A., Brezitskaya, V.V., Nikiforov, A.Y. and Smirnov, N.A. 2016. *Combined string searching algorithm based on knuth-morris-pratt and boyer-moore algorithm*s. IOP Conference Series: Materials Science and Engineering (Vol. 122, No. 1, p. 012034).

[6] Aygün, S., Güneş, E.O. and Kouhalvandi, L. 2016. Python based parallel application of Knuth-Morris-Pratt algorithm. *Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 2016 IEEE 4th Workshop on (pp. 1-5). IEEE.

[7] Sugiyono. 2011. *Metode Penelitian Kuantitatif kualitatif dan R&D*. Bandung: Alfabeta.

[8] Pressman, Roger S. 2010. *Rekayasa Perangkat Lunak: Pendekatan Praktisi (Buku I)* Penerbit Andi, Yogyakarta.

[9] Kumar, M., Singh, S. K., & Dwivedi, R. K. 2015. A Comparative Study of Black Box Testing and White Box Testing Techniques. *International Journal of Advance Research in Computer Science and Management Studies*, Vol. *3*(10).

[10] Sedgewick, R., & Wayne, K. 2015. *Algorithms, (Deluxe): Book and 24-Part Lecture Series*. Addison-Wesley Professional.

[11] Levitin, A., & Mukherjee, S. 2003. *Introduction to the design & analysis of algorithms* (p. 576). Reading: Addison-Wesley.