International Journal of Informatics, Information System and Computer Engineering

Reengineering of Proxy Logging Monitoring System in XYZ Institution

Dian Dharmayanti, Adam Mukharil Bachtiar, F Ahmad

Informatics Engineering Department, Universitas Komputer Indonesia, Jalan Dipatiukur Nomor 112-116, Bandung, Jawa Barat, Indonesia
*Corresponding Email: dian.dharmayanti@email.unikom.ac.id

ABSTRACTS

This research was conducted to assist the head of the IT Operational & Network Education Division of the XYZ institution to reengineering the internet usage monitoring system. This monitoring system is reengineered so that query performance becomes better and the information presented is easily understood by the user. Reengineering the old system will produce a system model that is more in line with the requirements of the current system. The steps taken in this study, analysis of the monitoring system case domain, data analysis, data visualization reengineering, internet use monitoring system reengineering, software implementation and testing. In addition, testing the performance of the query system and usability testing for users from the new visualization results as the output of this study. Reengineering the internet usage monitoring system is a solution to improve query performance and make it easier for users to understand the information presented.

ARTICLE INFO

Article History:
Received 17 Sept 2023
Revised 01 Des 2023
Accepted 19 Des 2023
Available online 29 Feb 2024

Keywords:
Reengineering,
Loggin Monitoring,
System,
Institution

1. INTRODUCTION

The rapid development of technology in the field of information has had a major impact on aspects of people's lives, one of which is the use of internet facilities. Internet facilities, apart from having a positive impact on many circles, especially educational institutions, also have a negative impact if not utilized properly (Huda et al., 2016). Monitoring the use of the internet to make usage regulations can be a solution to reduce the negative impact on internet usage (Pontes et al., 2015).

XYZ institution has a system to monitor internet usage. This system has several functions, one of which is query execution function to display traffic usage information on a particular site. However, the query execution time takes a long time. Query execution time is due to inefficient reading and recording of user log data. Within a period of 1 day data reading was repeated from the first data every time data parsing. In addition, the site url accessed by the user has not been summarized based on sub domains so that the same domain accessed by the same user will be considered as a different site domain. In addition to query performance problems, the need for visualization to convey information to users does not meet user needs because users need a long time to understand what information to convey from the results of the visualization displayed as well as the limitations of the current system (Majthoub et al., 2018).

The purpose of this research is to help the query performance of the internet usage monitoring system to make it better and to make it easier for users to understand the information displayed on the visualization. Reengineering carried out on the system includes changes to the content of procedures / methods, database structure and the results of data visualization in the system. The results of this study can improve query performance in the internet usage monitoring system in the XYZ institution network and make it easier for users to get information from the results of visualization.

2. METHODOLOGY

The methodology used in this study has several stages which are shown in Fig. 1 (Roth, 2004).

Fig. 1. Research Methodology

Based on Figure 1, this research begins with the case domain analysis stage to determine problems that occur in the system. After that, continued with data analysis on the system. The next stage is data visualization reengineering to make it easier for users to understand information. Reengineering of the internet usage monitoring system by making changes to the content of procedures / methods so that query performance can be better. Then the system implementation stage is carried out to measure whether the system is functioning as needed. The last stage taken is testing the software to find out the system is functioning properly and knowing the level of user understanding of the visualization of the information presented.

3. RESULTS AND DISCUSSION

In this section, we will explain in detail each stage and research of the research.

3.1. Case Domain Analysis of The Monitoring System

The first stage that is done is the case domain analysis phase to find out how the system works in general with problems experienced by the system in more detail in order to find a solution. The problems that occur in the system are described in the following points.

a. Data reading and recording are executed again from data 1 to the last data. Not only the last data executed every time it parses data.
b. The site URL accessed by the user has not been summarized into a sub domain.
c. Data storage procedures / methods are not yet as needed.
d. Users find it difficult to know what information is cold conveyed from the results of the visualization.
e. Users take a long time to get information from the visualization results

3.2. Data Analysis

At this stage, an analysis of the database is carried out to determine whether the current data model is in accordance with the system requirements or not. The data model is analyzed in the form of a relationship scheme or creating a new data model that fits the system's requirements. The relation scheme used in the current system can be seen in Fig. 2.

An analysis was carried out and it was found that redundancies were found between the site_traffic table and the traffic table, which had several of the same attributes for the same information needs as well. This can be seen in Figure 2. Therefore, in the current system relation scheme, one table is deleted between the two tables. The site_traffic table is removed because it produces the least amount of information. Based on the problems that occur in the current system relation scheme, an analysis is carried out to create a suitable new data model. The new data model designed can be seen in the Fig. 3.

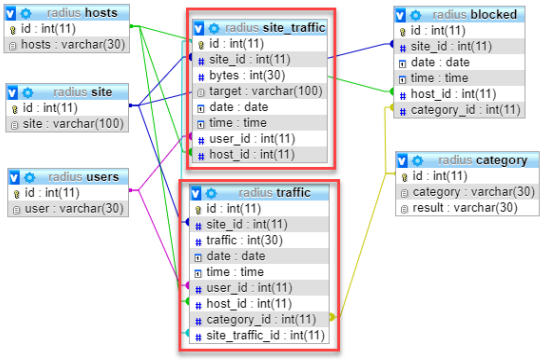

Fig. 2. Current Relationship Scheme

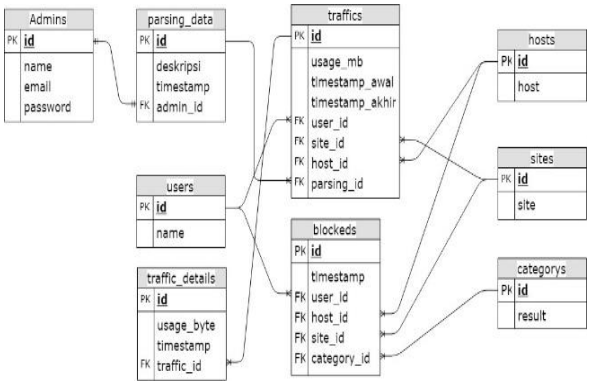

Fig. 3. Relationship Schema in the New System

Furthermore, to ease system performance in accessing queries or performing database-related processes, we must improve query performance (Zamzami & Budiardjo, 2013). Improvements in database performance that have been carried out during database creation are:

1. Selection of data types such as using enum data types to store string data (male and female), datetime to store date and time data rather than date / time data types, and avoid defining table columns with Null values.
2. The use of indexing on the appropriate data types such as unique for data strings, which is only one, index for foreign key data, and full text for string data in performing searches.

3.3. Reengineering Data Visualization

Visualization is a method that is often used to understand data effectively and efficiently. Effective data visualization will help users perform analysis and make complex data easier to understand (Fry, 2008). Before carrying out the data visualization process, it is necessary to know what information needs the user needs. The information needs based on users in the internet network monitoring process can be seen in Table 1.

Table 1. Information Needs Based on Users

User	Need
IT Operations at XYZ Institution	1. How to find out traffic usage in operating hours 2. How to find out 10 users with the most usage 3. How to find out Total Traffic Usage by date 4. How to find out the details of using Site Based Traffic

Creating an appropriate form of data visualization is necessary for understanding data. The process of converting data into information can be done in several stages (Bachtiar et al., 2018; Kandogan & Interfaces, 2012). The stages are carried out as follows :

1. Aquire
   Data source analysis (aquire) is done to find out what data sources are needed to support the needs based on the user. The data source used can be from several tables in the database.
   a. Table users
   b. Table sites
   c. Table hosts
   d. Table traffic
   e. Table traffic_details

a. Parse & Filter

At this stage the parse and filter process are carried out. Parse is the process of grouping data based on user needs, while filtering is the process of selecting appropriate data from the results of the parse process. The results of filter data mapping based on user needs can be seen in Table 2.

Table 2. Mapping Filter Data based on User Needs

Need	Table (parse)	Table Attribute (filter)
How to find out traffic usage in operating hours	users	name
	sites	site
	traffic_details	usage_byte, timestamp
How to find out 10 users with the most traffic usage	users	name
	traffic_details	usage_byte, timestamp
How to find out Total Traffic Usage by date	users	name
	sites	site
	traffic	usage_mb, timestamp_awal, timestamp_akhir
How to find out the details of using Site Based Traffic.	sites	site
	users	name
	traffic_details	usage_byte, timestamp

At this stage the represent and refine process is carried out. Representation is the process of changing data into an appropriate basic visualization model, while refining is a process of improving

the visualization model to make it more interesting and easier to read. The results of the visualization mapping can be seen in Table 3.

Table 3. Mapping Visualization

Need	Chart
How to find out traffic usage in operating hours	Line
How to find out 10 users with the most traffic usage	Bar Horizontal
How to find out Total Traffic Usage by date	Bar Column
How to find out the details of using Site Based Traffic.	List Tabel

a. How to find out traffic usage in operating hours

In this information needs used line chart visualization model. The choice of visualization model is because it is suitable for displaying the type of time periodical information. The results of visualization on this need can be seen in Figure 4.

b. How to find out 10 users with the most traffic usage

In this information needs used horizontal bar chart visualization model. The choice of visualization model is because it is suitable for displaying the type of cross section information (a description of the situation at a certain time). The results of visualization on this need can be seen in Figure 5.

c. How to know the total usage of traffic by date

In this need used bar chart column visualization model. The choice of visualization model is because it is suitable for displaying periodic information. The results of visualization on this need can be seen in Figure 6.

d. How to find out the details of traffic usage based on the site

In this need, the table list visualization model is used. The choice of visualization model is because it is suitable for displaying cross section information (an overview of the current state). The results of visualization on this need can be seen in Figs. 4-7.

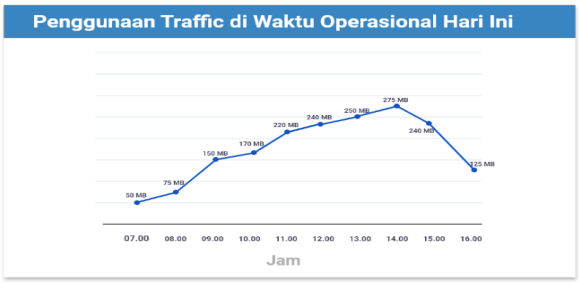

Fig. 4. Use of Traffic during Operations

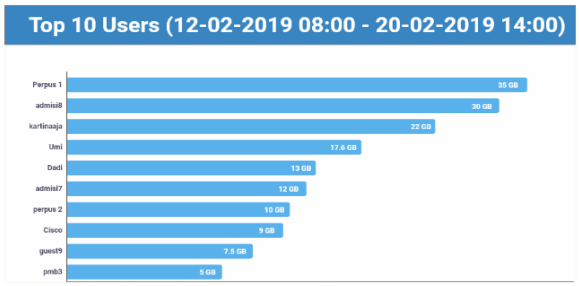

Fig. 5. Users With the Most Traffic Usage

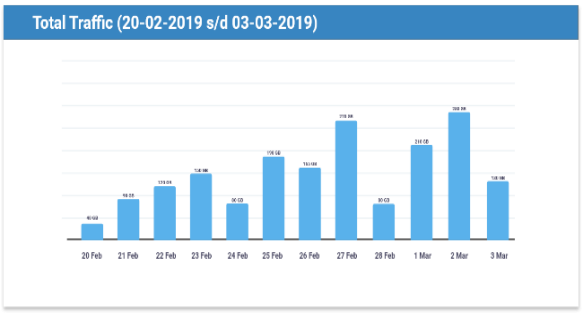

Fig. 6. Total Traffic Usage by Date

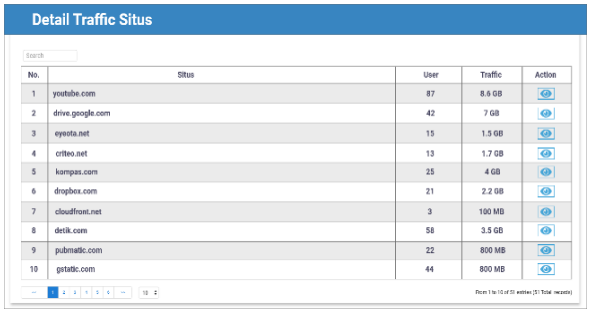

Fig. 7. Details of Traffic Usage by Site

3.4. Reengineering Monitoring System of Internet Usage

The approach used in this stage, namely reverse engineering, includes

analysis of the system subject in order to identify related system components and make representations of new system forms (Koschke, 2003). The first step is to analyze the functional requirements of the old system. This analysis aims to determine what procedures or functions the system has (Dipina et al., 2016). The functional requirements analyzed can be seen in Figs. 8 and 9. After analyzing the problems that cause the performance of the query to take longer, it is found that weaknesses in the methods and procedures used in the parse_data class can be seen in Table 4.

Table 4. Weaknesses of methods or procedures in the parsing process

Class	Method or Procedure	Information
parse_data	parsing_data()	Each time the data parsing process is executed again from data 1, it is not the most recent data that is executed
		url site that has not been summarized based on sub domains
	simpanData()	Source code for data storage conditions is not yet as needed.

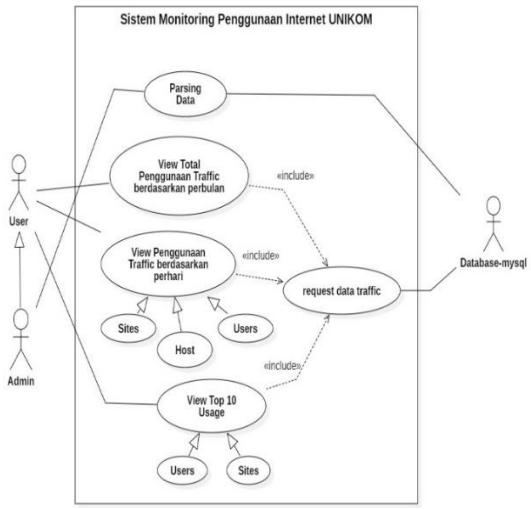

Fig. 8. Use Case Diagram of the current system

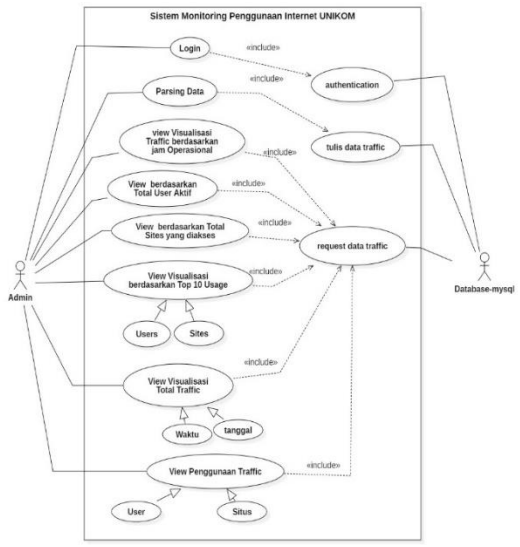

Fig. 9. Use Case Diagram of New System

For a comparison of the source code in the parse_data class in the current system with the new system, see Figs. 10 and 11. In the parse_data class in the parsing_data method, there is no check first on the last data that was saved to the database or executed. So that the user activity data that has been previously stored will be saved back to the database which results in the database becoming wasteful (Baron et al., 2008). Meanwhile, In the parse_data class on the new system called ParsingController, the parsing_data method is added to the source code for reading the last data that has been parsed and stored in the database so that the data to be stored in the database is not redundant.

```php
<?php
class parsing_data {

  function parsing_data (){
    $squidlog = "../../access.log";
    $file_squid = fopen($squidlog, "r");
    $parsedLogList = [];
    $i = 1;
    while ( (!feof($file_squid)) && ($squidlog != '') )
    {
      $buff_squid = fgets($file_squid);
      $array_squid = preg_split("/\s+/", $buff_squid);

      $log =[
        'date' => date('Y-m-d', $array_squid[0]),
        'time' => date('H:i:s', $array_squid[0]),
        'host' => $array_squid[2],
        'result_code' => explode('/', $array_squid[3])[0],
        'bytes' => $array_squid[4],
        'user' => substr($array_squid[7], 0, 50),
        'destination' => $squid_url = parse_url($array_squid[6])['host']
      ];
      array_push($parsedLogList, $log);
      $i++;
    }
    fclose($file_squid);
    return json_encode($parsedLogList);
  }

  //...

  return response()->json('success');
}

?>
```

Fig. 10. Data Parsing Method in the Current System

```php
<?php

namespace App\Http\Controllers;

//...

class ParsingController extends Controller
{
    //....

    public function parsing_data(){

      $endData = traffic::join('users', 'traffics.user_id', 'users.id')
                  ->join('sites', 'traffics.site_id', 'sites.id')
                  ->select('users.name','sites.site','traffics.timestamp_akhir')
                  ->orderBy('traffics.timestamp_akhir', 'desc')
                  ->first();

      $parsedLogList = [];
      $i=0;
      while ((!feof($file_squid)) && ($squidlog != '')) {

        if( $endData['timestamp_akhir'] <= $datetime &&
            ($endData['name']!=$user || $endData['site']!=$site) ){
          $log =[
            'no'=>number_format($i,0,',','.'),
            'dateTime' => $datetime,
            'host' => $array_squid[2],
            'result_TCP' => explode('/', $array_squid[3])[0],
            'bytes' => $array_squid[4],
            'user' => $user,
            'destination' => $site
          ];
          array_push($parsedLogList, $log);
        }

        $line = fgets($file_squid);
        $i++;
      }
      return response()->json($parsedLogList);
    }

    }
}
```

Fig. 10. Data Parsing Method in the New System

3.5. System Implementation

At this stage the implementation is carried out to the system from the results of reengineering that have been done. The implementation includes data and interface implementation. The data implementation includes tables of admins, users, sites, hosts, parsing_datas, traffic, traffic_details, blocked and category. Meanwhile, the interface implementation includes login, dashboard, Top 10 Usages by User, Top 10 Usages by Site, Total Usage by Date, Total Usage based on Time, User Traffic Details, and Site Traffic Details.

3.6. Software Testing

The testing phase is the last stage in this research. This stage is carried out to determine whether the functionality of the system is in accordance with the system's goals (Mifsud, 2005). The stages of system testing are divided into 2 namely performance query testing and visualization testing to users with usability testing.

a. Performance Query Testing

Query testing is performed to determine whether performance queries that are run on the new system will be better than performance queries on the old system. The query test results can be seen in Table 5.

Table 5. Comparison of Performance Queries

Need	System Query Performance	
	Current	New
Know the Total Traffic Usage on the "googlevideo" site category	4 minutes, 41.81 sec	0.21 sec

Based on Table 5, it can be concluded that the query performance that is run on the new system has increased. Improved query performance occurs due to changes in database structure and the addition of indexing to filed in certain data types such as indexes and full text.

b. Visualization Testing with Usability Testing

Usability testing is conducted to determine whether the visualization presented can facilitate the user in obtaining information. This test is done by giving assignments to users to complete. The assignment given to the user can be seen in the following scenario.

Table 6. Tasks in Scenario

No	Scenario	Task
1.	You are the Head of The IT Operations at XYZ Institution and want to find out information on traffic usage during operational hours.	1. What time is the highest traffic usage? 2. How much traffic is used for question number one? 3. What information is received? 4. What are the difficulties experienced in reading the results of the visualization presented? 5. Are there any suggestions on visualization results?
2.	You are the Head of The IT Operations at XYZ Institution and want to find out the information of 10 users with the most traffic usage.	1. Who has the highest traffic? 2. How much traffic is used for question number one? 3. What information is received? 4. What are the difficulties experienced in reading the results of the visualization presented? 5. Are there any suggestions on visualization results?
3.	You are the Head of The IT Operations at XYZ Institution and want to find out information on total traffic usage on each date	1. What date was the highest traffic usage? 2. How much traffic is used for question number one? 3. What information is received? 4. What are the difficulties experienced in reading the results of the visualization presented? 5. Are there any suggestions on visualization results?
4.	You are the Head of The IT Operations at XYZ Institution and want to find detailed information on traffic usage based on the site	1. How many users and traffic are accessed by the site at number 3? 2. How much traffic is used for question number one? 3. What information is received? 4. What are the difficulties experienced in reading the results of the visualization presented? 5. Are there any suggestions on visualization results?

The success of the test can be seen based on the three levels of assessment which can be seen in Table 7.

Table 7. The Three Levels of Assessment

Percent	Information
0%	Do not understand the scenario and not done at all
50%	Not completing the task perfectly or takes a long time
100%	Finish quickly and the goal is reached

The following are the results of tests conducted on users that can be seen in Table 8.

Table 8. Evaluation of Usability Testing

Participant	Scenario	Level of success	Time
Head of The IT Operations at XYZ Institution	1st	100%	6 sec
	2nd	100%	10 sec
	3rd	100%	13 sec
	4th	100%	8 sec

Results of the test will show success with views of efficiency and effectiveness. Efficient each task obtained from the task completion time of each participant, while effectiveness is seen from the success of the task. Efficient value and effectiveness of each task is 100%. It states that the purpose of usability testing is achieved.

4. CONCLUSION

Based on the results of the implementation and testing that has been done, it is concluded that the research reengineering the internet usage monitoring system can improve query performance. In addition, this system makes it easier for users to read the results of the visualization presented.

ACKNOWLEDGMENTS

The author would like to thank XYZ Institution for providing data and information relating to the object of research and UNIKOM as providing facilities in completing this research.

REFERENCES

Bachtiar, A. M., Dharmayanti, D., & Imammulloh, M. (2018, August). Data visualization of environmental factors in poultry farm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 407, No. 1, p. 012120). IOP Publishing.

Baron, S., Peter, Z., Vadim, T., Jeremy D, Z., Arjen, L., & Derek J, B. (2008). High Performance MySQL.

Dipina Damodaran, B., Salim, S., & Vargese, S. M. (2016). Performance evaluation of MySQL and MongoDB databases. *Int. J. Cybern. Inform.(IJCI)*, *5*.

Fry, B. (2008). *Visualizing data*. " O'Reilly Media, Inc.".

Huda, M. Q., & Hussin, H. (2016, October). Evaluation model of Information Technology innovation effectiveness case of higher education institutions in Indonesia. In *2016 International Conference on Informatics and Computing (ICIC)* (pp. 221-226). IEEE.

Kandogan, E. (2012, October). Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 73-82). IEEE.

Koschke, R. (2003). Software visualization in software maintenance, reverse engineering, and re-engineering: a research survey. *Journal of Software Maintenance and Evolution: Research and Practice*, *15*(2), 87-109.

Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.

Majthoub, M., Qutqut, M. H., & Odeh, Y. (2018, July). Software re-engineering: An overview. In *2018 8th International Conference on Computer Science and Information Technology (CSIT)* (pp. 266-270). IEEE.

Pontes, H. M., Szabo, A., & Griffiths, M. D. (2015). The impact of Internet-based specific activities on the perceptions of Internet addiction, quality of life, and excessive usage: A cross-sectional study. *Addictive Behaviors Reports*, *1*, 19-25.

Seffah, A., Donyaee, M., Kline, R. B., & Padda, H. K. (2006). Usability measurement and metrics: A consolidated model. *Software quality journal*, *14*, 159-178.

Zamzami, E. M., & Budiardjo, E. K. (2013). Sebuah Kajian Tentang Requirements Recovery Pada Area Riset Reverse Engineering. *JSI: Jurnal Sistem Informasi (E-Journal)*, *5*(2).