Development of Question Answering System for Public Relation Division in Student Admission

L R Maghfiroh[1], W Syahputra[1], I Santoso[1]

[1]Department of Statistical Computing, Politeknik Statistika STIS

*Corresponding author's e-mail: lutfirm@stis.ac.id

Abstract. Politeknik Statistika STIS (STIS) holds the new students' admission (PMB) every year which aims to gather, test, and, select all of its applicants who want to continue their study at STIS. STIS establish a committee during this event named Public Relation (PR) Division. PR Division to be intermediaries between STIS and the applicants. One of many PR Division tasks is to reply to all the questions from applicants about administration, procedure, or other things about PMB and STIS. PR Division is facing some problems that can hinder its performance to do the tasks. How do we address the problem is the reason that this research begins in the first place. The goal of this research is to build and establish a web-based system that is capable to solve all the problems the current system has. The system is divided into two main functions, the first one is FAQ management by PR Division members. The other function is a chatbot that automatically answers the question by using the TF-IDF algorithm. The conclusion on all testing and evaluation is the system that being build is already fulfilled all its requirements also the system is feasible to be used.

1. Introduction

Politeknik Statistika STIS (STIS) is a vocational university for the diploma programs managed by Statistics Indonesia (BPS). STIS has a vision to become an official higher education institution that functions to develop and disseminate knowledge, especially in the fields of statistics and statistical computing. [https://spmb.stis.ac.id/]

New Student Admission (PMB) is one of the annual activities that is routinely carried out by STIS. This activity aims to gather all prospective students who are interested in continuing their education at STIS from all regions in Indonesia. Since 2010, the registration process has been carried out online, and simultaneously in every province in Indonesia. After registering, the prospective student will continue according to the stages that have been determined previously.

In carrying out the PMB, STIS also established the Public Relations (PR) Division. PR Division has several tasks. One of them is to respond to incoming emails from registrants who ask about procedures and problems related to PMB registration.

Based on the subject matter presented, in 2017 PMB there were around 5000 emails sent to PR Division containing questions from the registrants. These emails have similarities in terms of the questions sent, such as questions about registration procedures, questions related to existing facilities at STIS, activities for the introduction and character-building (MP2K), as well as questions related to learning activities and service ties at STIS. These emails are also still being replied to manually by the PR Division, causing a lack of efficiency in handling them. Another problem is that there is no documentation of the problems that have been answered previously so that only the person who

handles the problem knows the problem. This results in differences in the knowledge of each member. In addition, to be able to answer emails, email access must be shared with all members. so that everyone can see all the emails. it is certainly not secure. In addition, shared emails result in ineffective email management, which may result in missed emails.

In 2018 the number of emails also increased. In 2019 and 2020, the handling of inquiries is served by a helpdesk system. However, this system is only to make managing email easier, while the email response process is still manual even though the questions asked are the same. The number of incoming emails will not decrease, especially in 2020, and still in 2021, Indonesia is hit by the COVID-19 pandemic, which causes all activities to be carried out remotely or online. The socialization process related to PMB, which is usually done with offline visits, cannot be carried out. So that the question and answer process is served entirely online.

According to Sneiders [1], to increase the level of efficiency in handling the incoming questions, the help desk FAQ Answering system can be used. The system can also be combined with a chatbot whose job is to reply to questions related to the implementation of new student admissions activities automatically. This system can be applied at PMB of STIS. The system works if there is an incoming question related to PMB, the question will be identified according to the question category, then answered according to the FAQ library that has been created previously.

Sudyana [2] built a help desk FAQ Searcher system for Consultation in Statistics Indonesia. The research contains a design system for the help desk FAQ Searcher in assisting the consultation process at BPS RI. The system uses a search function and TF-IDF [3] in matching the questions entered by the asker into the existing FAQ library. The difference between this system and the system to be built is the system that will be built using a chatbot in the question and answer process so that it is more interactive.

Yafie [4], the use of chatbots as an automatic document finder of Statistics Indonesia. In this study, the chatbot system was used as an automatic search engine for the results of BPS documentation that can be used by the public. The difference with the system that will be built is that this system is used to search for pdf documents, while the system that will be built uses the chatbot feature to search for FAQs.

Thus to deal with problems in the existing system, we need a system that can handle incoming questions related to PMB STIS automatically, as well as a knowledge management system for questions and answers related to PMB STIS. Therefore, the main purpose of this research is to develop the existing STIS PMB Public Relations Question and Answer system to be able to carry out their duties more effectively and efficiently.

2. Method
The framework of this research can be seen in Figure 1.

2.1. Data Collection Method
In this study, there are four methods used in the data collection process, namely:

2.1.1. Interview. In conducting system analysis, we used the interview method in doing fact-finding of the related system. Interviews were conducted on the subject matter related to the current system, in this case, the employees and lecturers who had been elected to the PR Division of PMB STIS. This interview was conducted so that an overview of the process of asking and answering questions in the PR Division, as well as what needs are needed in developing a proposal system.

2.1.2. Information Extraction FAQ. FAQ Information Extraction is done when there are new questions that come in from the asker. These questions will be answered by the Public Relations section of the PMB STIS who is in charge, then the questions and answers will be included in the FAQ library so that if the same question is entered, it will not need to be answered manually again by the PR Division of PMB STIS.

2.1.3. Survey. The survey was conducted to collect data related to the System Usability Scale (SUS) [5]. The results of this survey will later be used as an indicator of the usability of the system built for use by users. So that the results of this survey will be used as an analytical tool in this study.
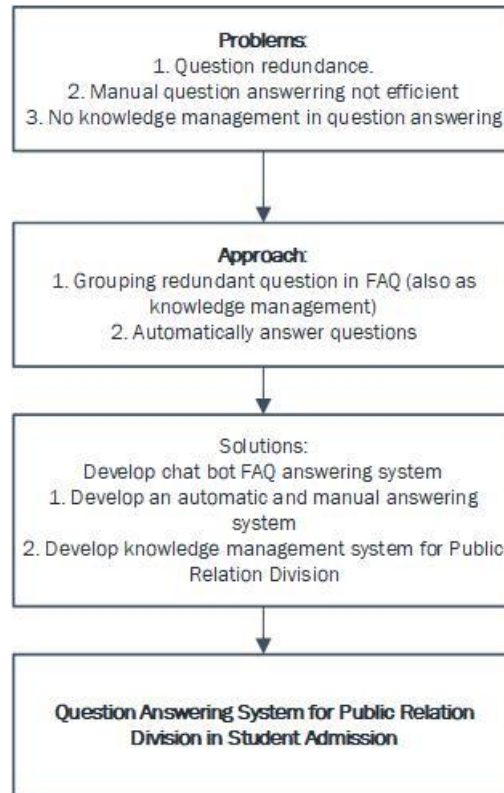

Figure 1. Research Framework

2.1.4. Literature Study. The literature study was conducted to collect related theories and research similar to this research. Literature studies are carried out by reading books, journals, and other studies so that they can be used to help complete our research.

2.2. Analysis Method.

Based on the results of collecting data related to system development, it has been successfully obtained that the system's business processes are currently running. Based on the exposure of subject matter related to the business process of PMB STIS Public Relations in answering questions sent by prospective students, it is illustrated in the business process diagram in Figure 2.
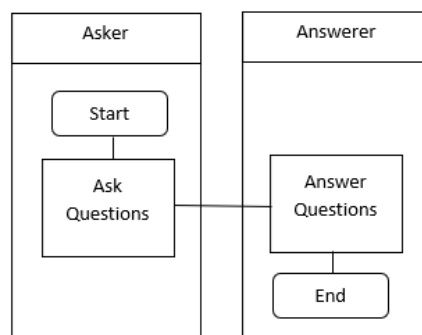
Figure 2. Current Business Process

Based on the business process in Figure 2, it can be seen that in carrying out the task related to answering incoming questions, the questions sent by the asker will be answered by the answerer via email. The incoming questions are in the same email, thus requiring all PR members to access one email at the same time.

In addition to business processes, the results of data collection also produce a set of problems that are still faced by the system that is running. These problems are summarized and illustrated in a fishbone diagram [6] Figure 3.

Based on the results of the previous analysis, the proposed system has several needs that must be met. These needs can be divided into two categories, namely functional and non-functional requirements:

a. Functional Requirements
   1. The new system is expected to be able to answer questions quickly and correctly.
   2. The new system should be able to automatically reply to duplicate questions and FAQs from many incoming questions.
   3. The system must also be able to accept questions that do not have an answer in the FAQ and be able to provide answers to these questions to the asker if they have been answered by a member of Public Relations.
   4. If there are questions that are not yet in the answer library, the system is also expected to be able to update the question library in accordance with the answers given by the Public Relations Section.

b. Non-Functional Needs
   1. The system is expected to be active for 24 hours.
   2. The system must have a security system that can distinguish between inquiries by humans and from spam bots
   3. The system must be able to provide the data needed by the user exactly when needed.
   4. The system must have an effective response time.
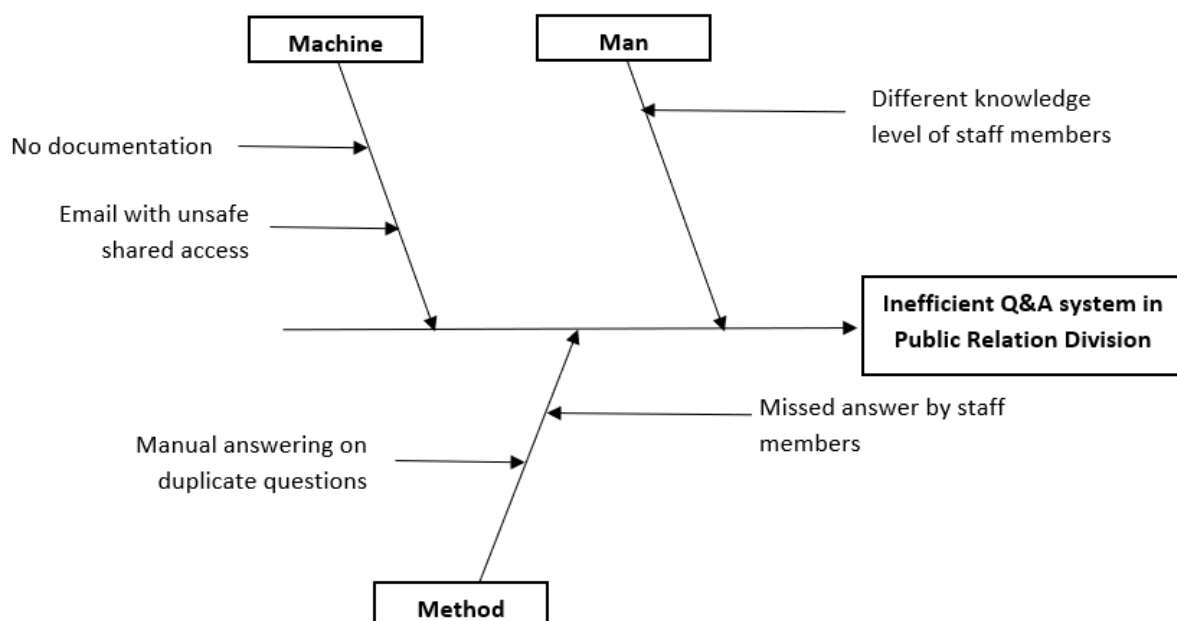   5. The output provided by the system must match the input entered by the user.

Figure 3. Fishbone Diagram of Current System

2.3. Evaluation Method

In evaluating the system, researchers used four different methods. Black box test [7] were carried out to evaluate the features of the system. The system usability scale (SUS) test was conducted to evaluate

the user's perception of the system being built. The Precision and Recall test [8] was conducted to evaluate the accuracy of the chatbot in identifying questions.

3. Result and Evaluation

3.1. Proposed System Design

In the system design, there are three main actors in the process of asking and answering questions related to PMB STIS, and there are additional actors that run automatically, namely chatbots that function as intermediaries in asking and answering activities. To understand the tasks of each actor, as well as to understand the business process of the design system, it can be seen in the use case diagram [9] and the business process in Figure 4 and Figure 5.

Figure 4 shows there are three actors. First, the asker needs to verify himself by entering data in the form of name, email address, and captcha if he wants to ask a question that has not been included in the previous database. Meanwhile, if you want to ask a question and see the answer, the asker can do this directly on the main page. The next actor is a member of public relations. Public relations members are required to log in first. After logging in, public relations members have the authority to answer questions that have just been entered by the asker that does not have answers in the database. Answers given by PR members will be stored first for verification by the super admin before being sent to the asker. The last actor is the super admin, this actor has the authority to manage the FAQ database. Things that can be done are adding, deleting, or changing FAQ questions and answers contained in the database. In addition, the super-admin also has the right to manage the member database, namely the right to add or remove PR members from the system, as well as verify answers from PR members.
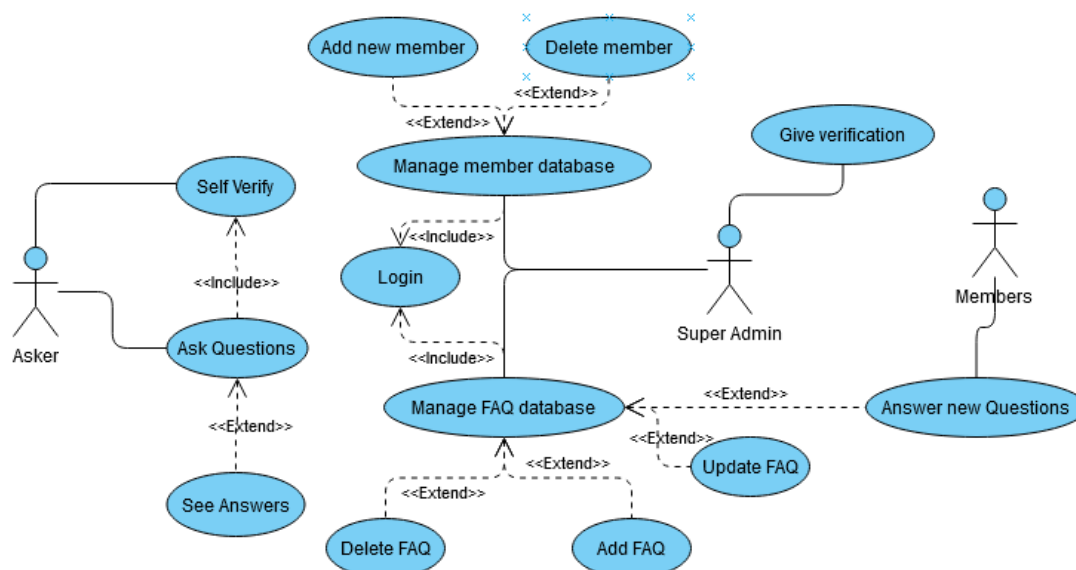

Figure 4. Use Case Diagram

In Figure 5, it can be seen that the flow of the system starts from the asker who writes questions for the chatbot. Then the chatbot will look for whether the question is already in the FAQ database or not. If the question is already in the FAQ database, the chatbot will immediately display the answer to the question. However, if the question asked by the asker is a new question that has not been included in the previous database, the chatbot will direct the asker to enter the question into the database. The asker is required to write down self-verification first in the form of the name and email of the asker, as well as a captcha code for security reasons. If the question has been answered by a PR member, the answer is verified first by the super admin. After that, the answer will be sent to the email of the asker directly.
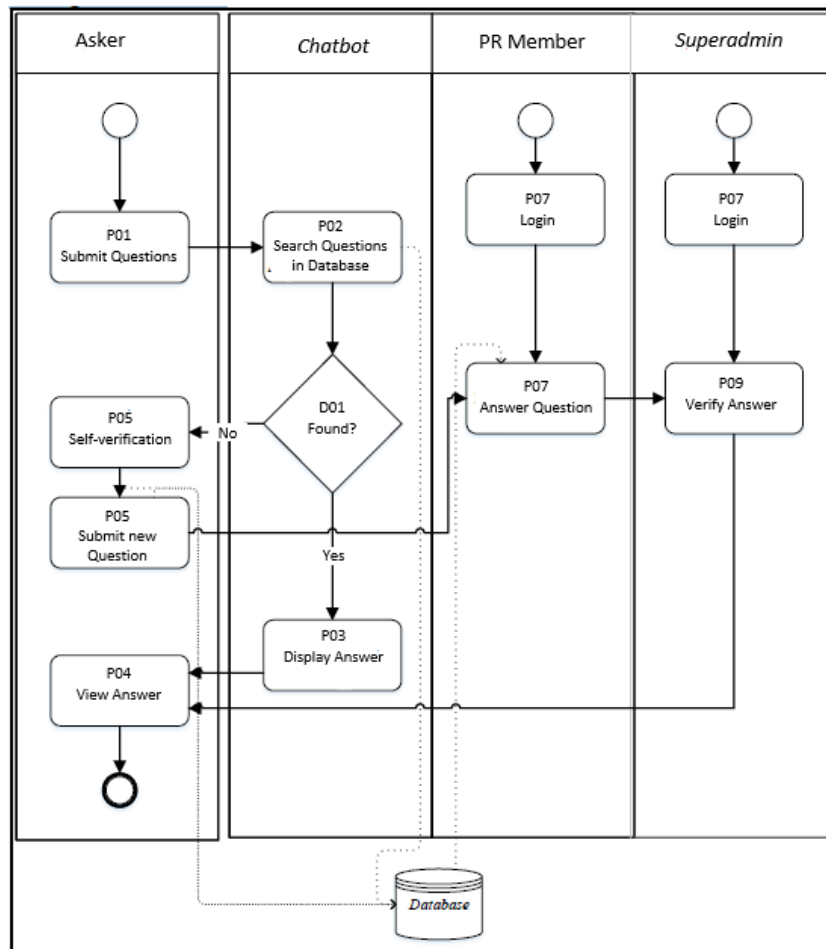
Figure 5. Proposed Business Process Diagram

3.2. Proposed System Implementation

The design system is divided into two parts, namely for the asker and the answerer. First, it will explain all the pages that will be used by the answerer who is a member of public relations. The first page that will be accessed by PR members is the login page in Figure 6. After the PR members successfully log in (Figure 6), the user will be redirected to the main dashboard page of the PMB STIS PR helpdesk system. With the display in Figure 7.

Figure 6. Login Display

Figure 7. Dashboard

The next page is a login question. This page will display all questions submitted by the asker manually. There are two options for each of these questions, namely, answer and delete. The delete option will automatically remove the question from the database. While the answer options will redirect the user to the answer page in Figure 8. The name and email fields cannot be changed by the user, but the question can be changed by the user according to their needs. Because when the submit button is selected. The system will save the questions and answers into the FAQ database automatically. In addition, the system will also send the questions and answers automatically to the asker's email address.

The last menu available is the member view menu. This menu will only appear if the logged-in user has super admin status. So if the one who is logging in does not have an admin user status, the menu will not appear on the site page. On the member view page (Figure 8) there is a table that contains all public relations member accounts contained in the database. There is a delete option to delete the related member data from the database. In addition, there is also a new add menu to add an account for new members.

Figure 8. Member view display page

For the asker, there is one main page containing a chat box (Figure 9). In the chatbox, the chatbot will display answers to questions entered by the user. The page also displays a question guide for users to make it easier for users to find answers.
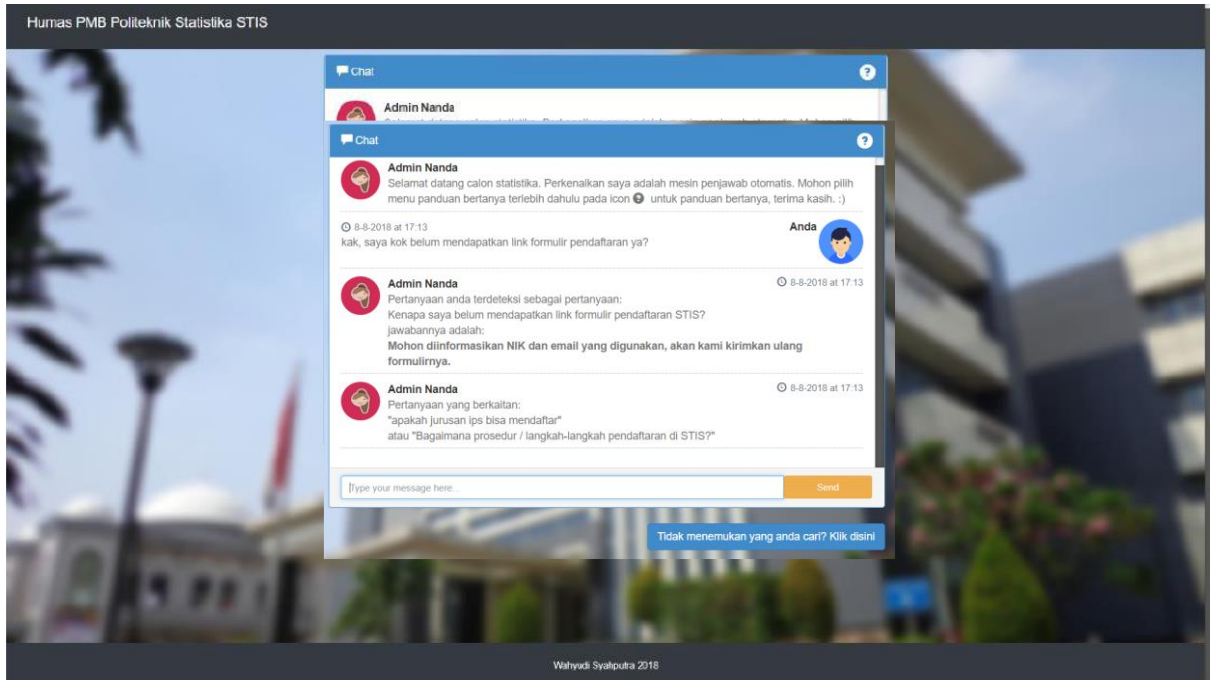

Figure 9. Chatbot page display

Users can directly type the question they want to ask in the ask column provided in the chatbox. When the user enters a question, the reply chat function contained in the program will be executed. In this function, four steps will be executed sequentially, namely stemming [10] question sentences, removing words that are included in the stopword list, checking for typos using the Levenshtein distance method [11], finally sorting the FAQ using the TF-IDF calculation method [12]. The stemming and stopword functions used by the author are stemming functions that come from the literary library on the packagist.org website [13]. The stemming and stopword functions used by the author are stemming functions that come from the literary library on the packagist.org website [13]. The stemming function uses the Nazief and Adriani algorithms [14] and can be directly called to cut punctuation marks and affixes to words from sentences entered by the user. The stopword function in the library has also automatically collected a list of all stopwords in the KBBI. After the question sentence goes through the stemming and stopword stages. Then tokenization will be carried out and check for typing errors first using the Levenshtein distance method [13]. In checking for typing errors, the author uses the typing error checking algorithm from the study [13]. The author chooses the limiting value in the Levenshtein distance algorithm which is in line with [2].

3.3. Evaluation

The SUS test is carried out by asking 15 respondents, who had helped with the PR Division task of PMB STIS, to run all the features contained in the system. This feature comes from the previous blackbox test so that it can be seen whether the features on the system are running properly. Based on the test results, the final average was obtained with a value of 85.5. Based on this value, it can be said that the system built is acceptable to the user.

In addition, during the implementation of the SUS, respondents were also asked to assist the writer in finding the accuracy level of the detection of questions owned by the chatbot by asking respondents to provide questions related to a topic contained in the FAQ, then respondents were asked to ask the same thing with sentence structure and vocabulary differently. To then calculate the percentage of

chatbot accuracy in detecting questions. For the precision and recall test, where the precision test results show a value of 90.667 percent (%), and the recall value is 100 percent (%).

Although the evaluation results show good results, the system can still be improved after being implemented for naive public users. Thus the value of the feasibility and accuracy of the results can be increased so that user satisfaction is further increased.

4. Conclusion

In this study, a question answering system was successfully built to help the Public Relation Division to help answering questions automatically if the question already had an answer stored in the system. In addition, through this system, the asker and the public relations division member can conduct manual question answering if the question does not have an answer stored in the system yet. Based on the evaluation of the system has been accepted by the user.

References

[1] Sneiders, Eriks. (1999). Automated FAQ Answering: Continued Experience with Shallow Language Understanding. AAAI Technical Report FS-99-02.

[2] Sudyana, A. (2015). Pembangunan Sistem Informasi Helpdesk Modul Sistem Tanya Jawab. Jakarta Timur: Sekolah Tinggi Ilmu Statistik.

[3] Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries. New Jersey: Department of Computer Science, Rutgers University

[4] Yafie, A. R. (2017). Pengembangan Sistem Penjawab Otomatis Berbasis Chat untuk Pencarian Data Publikasi Badan Pusat Statistik. Jakarta Timur: Sekolah Tinggi Ilmu Statistik

[5] Bangor, A., Kortum, P. T., & Miller, J. T., (2008). An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction.

[6] Illie, G & Ciocoiu, C. N. (2010). Application of Fishbone Diagram to Determine the Risk of an Event with Multiple Causes. Management Research and Practice Vol. 2 Issue 1 March.

[7] Everatt, G.D.; McLeod Jr., R. (2007). "Chapter 2: The Software Development Life Cycle". Software Testing: Testing Across the Entire Software Development Life Cycle. John Wiley & Sons. pp. 29–58. ISBN 9780470146347.

[8] Ting, K. M. (2011), Precision and Recall. Springer, Boston, MA

[9] Dennis, A., Wixom, B. H., & Roth, R. M. (2012). System Analysis and Design. Vol. Fifth Edition. USA: Wimbley

[10] Frakes, W.B. & Baeza-Yates, R. (1992). Information Retrieval – Data Structure and Algorithm. Prentice-Hall, Inc

[11] Navarro, Gonzalo (2001). A guided tour to approximate string matching. ACM Computing Surveys.

[12] Xia, T., & Chai, Y. (2011). An Improvement to TF-IDF: Term Distribution Based Term Weight Algorithm. Journal of Software Vol. 6 No. 3.

[13] Librian, A. Packagist Sastrawi. https://packagist.org/packages/sastrawi/sastrawi. diakses pada 19 Juli 2018.

[14] Agusta, L. (2009). Perbandingan Algoritma Stemming Porter dan Algoritma Nazief-Adriani Untuk Stemming Dokumen Bahasa Indonesia. Konferensi Nasional Sistem dan Informatika 2009.