Jurnal Sains dan Teknologi
Volume 11 Number 1, Tahun 2022, pp. 132-141
P-ISSN: 2303-3142 E-ISSN: 2548-8570
Open Access: https://dx.doi.org/10.23887/jst-undiksha.v11i1

Web-Based Chatbot with Natural Language Processing and Knuth-Morris-Pratt (Case Study: Universitas Internasional Batam)

Meiliverani Erline[1*], Yefta Christian[2]

[1,2] Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Internasional Batam, Indonesia

ARTICLE INFO

Article history:
Received January 06, 2022
Revised January 09, 2022
Accepted April 04, 2022
Available online April 25, 2022

Kata Kunci:
Chatbot, Pemrosesan Bahasa Alami, Knuth-Morris-Pratt.

Keywords:
Chatbot, Natural Language Processing, Knuth-Morris-Pratt.

ABSTRAK

Sebagai sebuah institusi pendidikan, akan banyak orang yang mencari informasi tentang Universitas Internasional Batam (UIB), baik itu mahasiswa, alumni, calon mahasiswa, rekan kerja, maupun masyarakat luas. Saat ini layanan informasi di UIB masih dilakukan secara manual oleh manusia atau melalui website. Pengguna perlu melakukan beberapa langkah dengan menggunakan fitur pencarian berbasis teks. Penelitian ini dilakukan untuk membangun sistem chatbot berbasis web dengan menerapkan pendekatan NLP dan dilengkapi dengan algoritma KMP. Penelitian ini menggunakan teknik web scraping untuk mengumpulkan dan mengekstrak data yang dibutuhkan dari situs resmi UIB sebelum diolah menggunakan pendekatan Natural Language Processing dan algoritma knuth-morris-pratt. Chatbot dibangun menggunakan bahasa pemrograman Python dengan harapan dapat menjawab pertanyaan dengan benar secara real-time dan memudahkan pengguna untuk mendapatkan informasi tanpa harus kuliah. Dataset pelatihan chatbot terdiri dari 193 pertanyaan dan pernyataan yang dikelompokkan ke dalam 18 tag. Sedangkan teknik Black Box Testing diimplementasikan untuk pengujian chatbot menggunakan 100 soal. Chatbot berbasis web ini berhasil dibangun dengan tingkat akurasi 86% dalam menjawab pertanyaan pengguna.

ABSTRACT

As an educational institution, many people will seek information about Batam International University (UIB), including students, alumni, prospective students, colleagues, and the wider community. Currently, information services at UIB are still done manually by humans or through websites where users need to take several steps using the text-based search feature. This research was conducted to build a web-based chatbot system by applying the NLP approach and using the KMP algorithm. This study uses a Web Scraping technique to collect and extract the required data from the official UIB website before processing it using the Natural Language Processing approach and the Knuth-Morris-Pratt algorithm. The chatbot was built using the Python programming language to correctly answer questions in real-time and make it easier for users to get information without having to go to lectures. The chatbot training dataset consists of 193 successful questions and statements in 18 tags. The Black Box Testing technique is implemented for testing chatbots using 100 questions. This web-based chatbot has been successfully built with an 86% accuracy in answering user questions.

1. NTRODUCTION

Today, technology plays an important role in society, especially in providing all forms of information (Dahiya, 2017; Danniels et al., 2020; Wahyuni et al., 2020). Most of the time, people have to visit colleges or its website to gain information about it (Eze et al., 2018; Mpungose, 2021; Tovar Viera et al., 2020). Visiting college is somehow tedious, it requires manpower and time just to provide the needed information. Visiting a website to search information through it is better than visiting college itself, but it is still less efficient because the user has to do the text-based searching (Manuhutu & Wattimena, 2019; Rasmila, 2018). The user still has to manually search for the required information, because one website page might be containing a lot of information (Wilujeng et al., 2020). One of the techniques that have been widely used to support the provision of information is artificial intelligence (AI) techniques (Kakani et al., 2020).

AI is widely developed to support and improve the quality of decision making and problem solving in various industries (Bertram et al., 2021; M. Sharma et al., 2021). Development is carried out through the exploitation of various types of machine intelligence, including natural language comprehension and machine learning (Androutsopoulou et al., 2019). In this rapidly evolving connected world, not only human beings can interact online. Most of the tech things have a bot to interact with people (Anghelescu &

*Corresponding author.
E-mail addresses: 1831126.meiliverani@uib.edu (Meiliverani Erline)

Nicolaescu, 2019). A technology that is designed to mimic the human conversation called Chatbot is one of the choices (Pi & Majid, 2020). Chatbot is a program designed to promote interactive conversation or communication to users through text, sound, visual media such as images and videos. Chatbots with artificial intelligence are able to make it easier for users to obtain information quickly and precisely (Wijaya et al., 2020).

Commonly, chatbots are developed using the natural language processing (NLP) approach which is part of the branch of AI (Cheng et al., 2022; Prianto & Harani, 2021). Veronika, Botika, and AiChat Indonesia are some of popular chatbots that have been developed using NLP and rule-based methods (Chandra & Suyanto, 2019). In order to improve the quality of a chatbot, an effective and efficient search technique is needed. Knuth-Morris-Pratt algorithm is one of the appropriate search techniques. The KMP algorithm works by matching the pattern per character from left to right one by one until one of the conditions is met. That's why the KMP algorithm is very compatible with all types of string searches. Search techniques with KMP have been used in many applications because they can minimize computing time, especially in big data (Ependi & Oktaviani, 2017; Kalnoor & Agarkhed, 2018). Through the Black Box testing technique, the study also concluded that in designing the abstract keyword search program of the student's final project, the search process with the KMP algorithm shows results that are 100% the same as the keywords entered.

Web scraping is used to scrape information from the website. Scraping technique is the process of retrieving a semi-structured document, then extracting it to get certain data from the page. By scraping, the data retrieved is more focused and easier to get (Arisandi et al., 2021; Hajikhani et al., 2022; Muehlethaler & Albert, 2021). There are some earlier studies about the implementation of chatbot with NLP. One of them is the application of Covid-19 chatbot with NLP approach by (Prianto & Harani, 2021; A. K. Sharma et al., 2021). There are also many earlier studies that implemented chatbot as college information management. UNIBOT is one of the examples, a web-based human-like interactive university chatbot (Grandi et al., 2011; Patel et al., 2019). This research is done with the aim of making it easier for users to get information about college quickly and easily. Hence, an automation system such as chatbot is developed. With the implementation of NLP approach to process colloquial language, and Knuth-Morris-Pratt algorithm to perform the pattern matching, to determine the most appropriate answer to the questions.

2. METHOD

This research is done to build a web-based chatbot system by implementing the NLP approach and complemented by KMP algorithm. Firstly, data collection is required to gather data to generate questions pattern and answers template. Web scraping technique is used to scrape data from UIB website with JSON file outcome. The JSON file is used to train the chatbot. The generated chatbot models from the training process is called every time the chatbot is operated and interacted. Then the knowledge of chatbot is compared to user's question with KMP. Before being compared, user's question is processed with NLP using NLTK for chatbot to be able to understand the query. Finally, the Flask framework is used to return the answers back to users. Hence, the first step is to develop Web Scraping system with Scrapy, then the chatbot training to generate the chatbot models, last is to develop the implementation of NLP, KMP, and Flask framework. All these steps are followed by testing the accuracy of the chatbot presented in Figure 1.

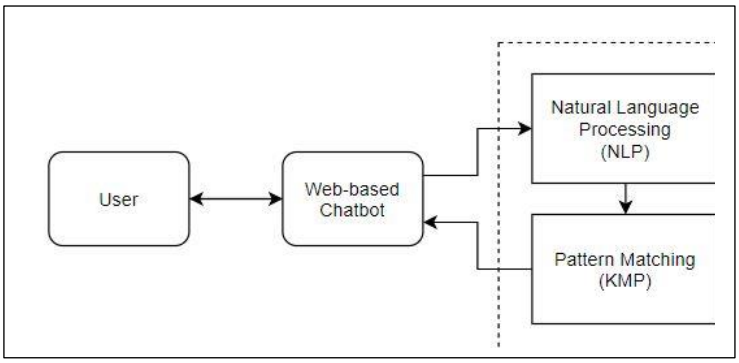

Figure 1. System General Overview

The explanations of the chatbot flow according to the general overview are presented as follows. First, users should visit the URL of the web-based chatbot to be able to start the chat without any account required. Second, Users could send messages and the built chatbot is responsible for getting the message to be processed. This step is done with the application of Flask App Routing. Third, The backend of the chatbot, which is written in Python, will tokenize and lemmatize the message using the Natural Language

Toolkit (NLTK) which is a toolkit in the form of go-to API for working with NLP in Python. The tokenization is used to remove white spaces, commas, and any other symbols in order to split the message into tokens (terms). It is an important step for extracting words and lemmas. While lemmatization is another NLP technique which function is to reduce any inflectional forms or derivational forms of a word into its basic form (Khader et al., 2018).

The pattern matching will be run using the KMP algorithm based on the knowledge pattern that has been created. The pattern is grouped into several tags. The extracted words and lemmas from user messages are going to be run in a loop to do a comparison with the pattern. Every word that matches the pattern will be stored as 1 in a matrix of N words with N is the number of words in the pattern. The trained model is used in this step to predict the answer by sorting out the strength of probability. The answer will be chosen randomly from the response template in a tag with the highest probability and the error threshold does not exceed 0.25. The appropriate answer chosen will be sent back to the user. Data availability and knowledge of end-user needs will help to ease the system design and development. This stage is conducted to collect all data as needed by web scraping technique. A free open-source web scraping framework named Scrapy is utilized in this research for extracting the data needed from UIB website. As seen in (Jabeen, 2019), a simple overview of how Scrapy works as a Python framework is illustrated in Figure 2. The data crawled from the web will be parsed and extracted into a JavaScript Object Notation (JSON) format with the structure.

```
{'intents': [
    {'tag': 'greeting',
  'patterns': ['Hallo!', 'Hi, how are you?'],
  'responses': ['Hello!', 'Hi there, how can I help?']
    }
]}
```

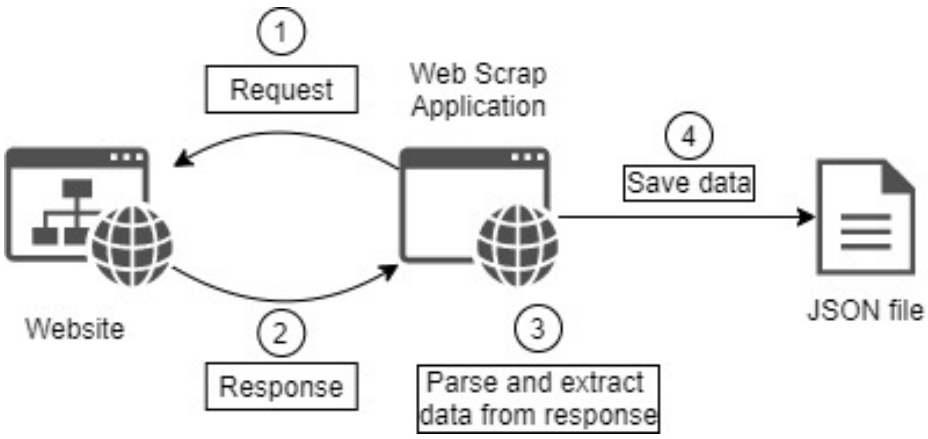

Figure 2. Web Scraping Overview

The use of artificial intelligence on this chatbot system lies in the pre-processing process, specifically using natural language processing. NLP is a branch of AI concerned with enabling computers to understand words in much the same way as human being (Surabhi, 2013). The pre-processing process includes tokenization and lemmatization. This method is applied to the process of preparing a bag of words for chatbot training based on the pattern and response file. This pattern and response file is in json and divided into several tags, such as greeting, thanks, goodbye, no answer, address, contact, news, event, announcement, scholarship, program, etc. The tag and pattern are obtained from summarizing all information that may be obtained from the website used for web scraping, in this case UIB website. Meanwhile the response is obtained from the result of web scraping. Thereafter, this method is applied for processing the user's response in advance of doing the pattern matching to find the answer.

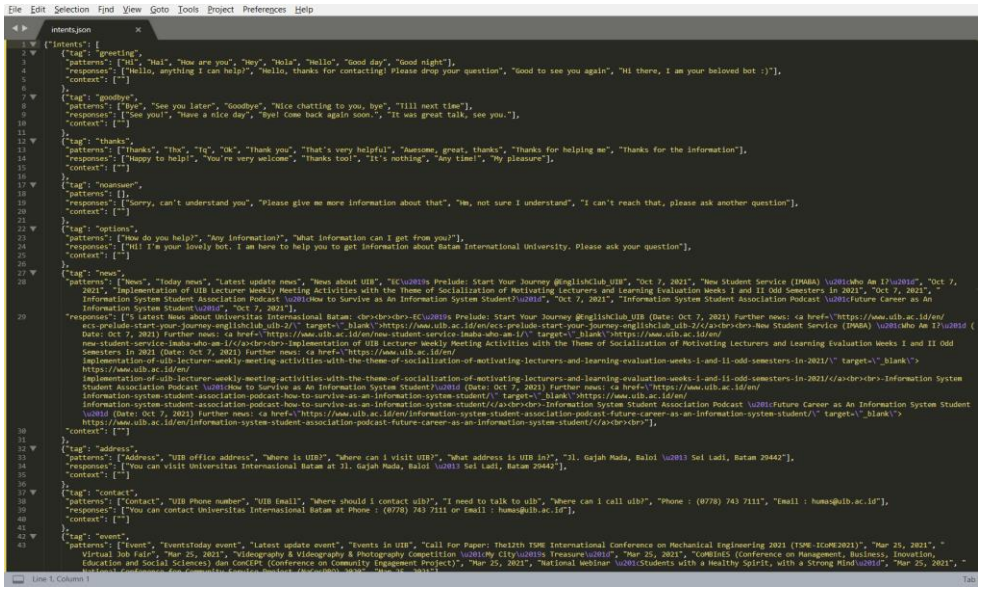

Figure 3. Pattern and Template JSON Output from Web Scraping

The use of Knuth-Morris-Pratt as a string search algorithm in this web-based chatbot lies in the pattern matching process. The pattern matching is being done based on the user message or response to the bot that has gone through the preprocessing process. This algorithm is one of the Exact String Matching Algorithm types, which is an exact string matching with the arrangement of characters in a matched string that has the number and sequence of characters in the same string. KMP performs matching one character by one character from left to right (Rahim et al., 2017). Prefix and suffix of the matching is used in the KMP algorithm. It's known as Longest Proper Prefix which is Suffix (LPS). LPS is basically a one-dimensional array consisting of a number of elements which is equal to the number of characters in the pattern (Dawood & Barakat, 2020). KMP algorithm consists of pre-processing and main stage. In the pre-processing stage, LPS array or failure table is generated for each word in order to identify the maximum possible shift to prevent the unnecessary comparisons while searching. This action could also decrease the needed time when comparing pattern and text string (Warnars et al., 2021). The implementation of KMP preprocessing stage by generating a failure table is shown in Table 1.

Table 1. Failure Table

Word	LPS Array
goodbye	[0, 0, 0, 0, 0, 0, 0]
scholarship	[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
AAACAAAA	[0, 1, 2, 0, 1, 2, 3, 3]

As the failure table has successfully generated, it will be used while matching the pattern and text string in order to decrease matching time. The main stage of KMP algorithm will be implemented. Both the pattern string and text string that wanted to be compared are required as an input in order to run the KMP algorithm function above. Table 4. to 7. are the visualization of string matching with KMP. Each table consists of text string and its index, pattern string and its index, and the LPS array of the pattern. The comparison of text "newer news" and pattern "news" started with variable i and j defined as 0 to perform shift calculation (Anggreani et al., 2020). Variable i represents text index while j represents pattern index. Index=0 of pattern (pat[0] = n) and index=0 of text (txt[0] = n) match as shown in Table 2.

Table 2. String Comparison with KMP 1st Comparison

Index	0	1	2	3	4	5	6	7	8	9
Text string	n	e	w	e	r		n	e	w	s
Pattern index	0	1	2	3						
Pattern string	n	e	w	s						
Pattern LPS	0	0	0	0						

Table 3. String Comparison with KMP 2nd Comparison

Index	0	1	2	3	4	5	6	7	8	9
Text string	n	e	w	e	r		n	e	w	s
Pattern index				0	1	2	3			
Pattern string				n	e	w	s			
Pattern LPS				0	0	0	0			

Table 4. String Comparison with KMP 3rd Comparison

Index	0	1	2	3	4	5	6	7	8	9
Text string	n	e	w	e	r		n	e	w	s
Pattern index					0	1	2	3		
Pattern string					n	e	w	s		
Pattern LPS					0	0	0	0		

Table 5. String Comparison with KMP 5th Comparison

Index	0	1	2	3	4	5	6	7	8	9
Text string	n	e	w	e	r		n	e	w	s
Pattern index							0	1	2	3
Pattern string							n	e	w	s
Pattern LPS							0	0	0	0

Since the first character match, comparison continues to the next character of text and pattern by adding the value of both variable i and j by 1. If a mismatch is found as shown in the 5th column of Table 2., the pattern will be moved to the last mismatch position (Warnars et al., 2021), which is in text index = 3. It's because the value of j is 3 at this stage, so variable j should be set to j = lps[j-1]. Value of pattern LPS with index = 2 is 0 as seen in the 7th column of the last row of Table 3. In this case, the LPS array is used to skip unnecessary comparison and it leads to the changing of comparison position as seen in Table 3. Mismatch occurs on the comparison of pattern index = 0 (pat[0] = n) and text index = 3 (txt[3] = e), so the value of j is considered (Dawood & Barakat, 2020). Since its current value is 0, the first character of pattern will be compared with the next character of text (txt[4] = r) as shown in Table 4. The comparison will continue looping character by character until the value of i reaches the last text index. All pattern characters match the substring in the text string starting from text index = 6 as shown in Table 5.

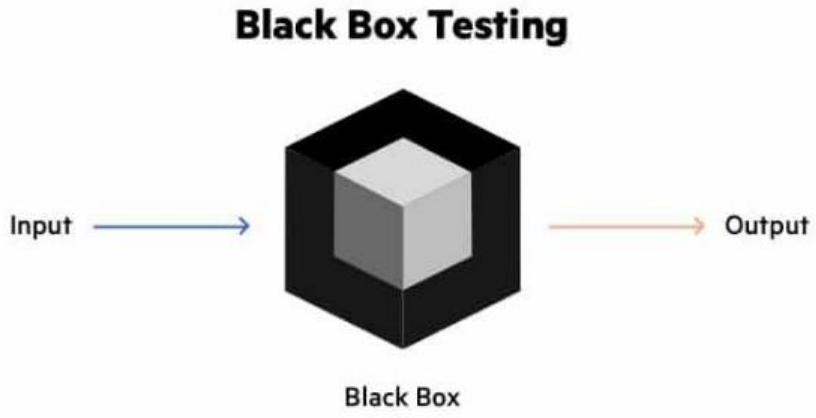

Figure 4. Black Box Testing

The stage of testing is a step where the User Acceptance Test (UAT) is conducted. Black Box Testing method is used in this study to determine whether the chatbot is running well according to the expectations or not. Black Box Testing only focuses on the accuracy of system functionality without looking at the structure and source code behind, in other words it ignores the internal mechanisms of the system (Supriyono, 2020).

3. RESULTS AND DISCUSSION

Result

Design of User Interface (UI) is such an important part of any system (Patel et al., 2019). This web-based chatbot is designed similar to chat application design in general. HTML, jQuery, CSS, and Flask are used to build the chatbot user interface. Front-end plays a big role in getting user messages, relaying it to the back-end, then showing the response back to the user. The design of the whole chatbot is shown in Figure 5.

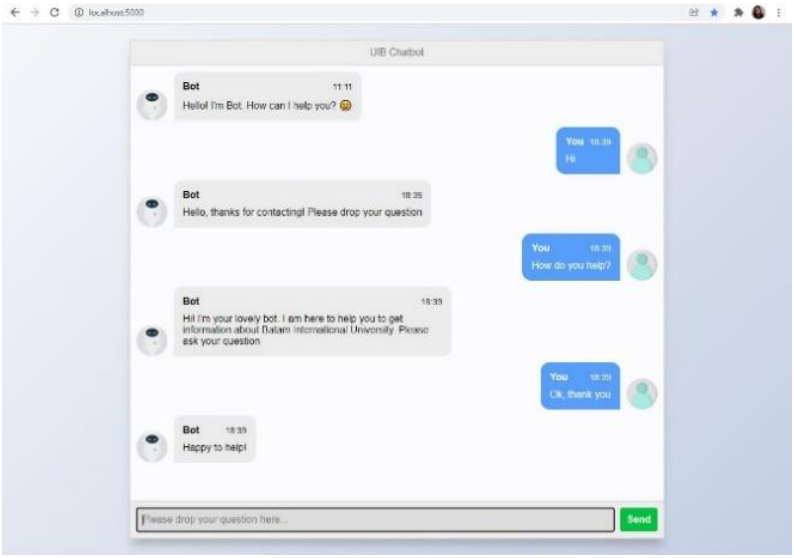

Figure 5. Chatbot User Interface

This research used 193 questions in 18 different intents which come from web scraping systems. The questions are manually created by the developer based on information from UIB's website. The web scraping system also generated answer templates from each intent. After those patterns and templates are trained and generated a chatbot model, the model is used each time the users throw a question. User's question is tokenized and lemmatized to make it easier for the bot to understand. The tokens and lemmas are compared to each question that have been trained to the bot. The comparison involves string matching with KMP. Next, the chatbot model predicted in which intent the question is. It's run to get one intent with highest strength of probability, which means the patterns of this intent match the user's question the most. After the intent is predicted, one response template is randomly chosen from this intent as the most appropriate answer. Users can operate the chatbot with following steps: First, users need to visit the chatbot URL. Second, users can immediately start asking the chatbot as soon as the GUI shows. Third, the chatbot will return the answer immediately to the users. This web-based chatbot is not integrated with UIB's website yet. So, users need to manually visit the chatbot URL. Integration with the university or college website is somehow recommended for next research since it helps the website visitors to reach the chatbot. Since the training data is limited by the patterns and templates JSON, the chatbot's knowledge to answer questions is limited. More data training and dynamic source of data will help to improve the chatbot ability to answer user's queries.

Table 6. Black Box Testing Result

No	User Input	Chatbot Response	Intent	Status
1	Hi, how do you help?	Hi! I'm your lovely bot. I am here to help you to get information about Batam International University. Please ask your question	Option	Success
2	I got it, thanks!	Happy to help!	Thanks	Success
3	See you later	Bye! Come back again soon.	Goodbye	Success
4	Where is UIB?	You can visit Universitas Internasional Batam at Jl. Gajah Mada, Baloi – Sei Ladi, Batam 29442	Address	Success
5	I need to contact UIB	You can contact Universitas Internasional Batam by Phone : (0778) 743 7111 or Email : humas@uib.ac.id	Contact	Success

No	User Input	Chatbot Response	Intent	Status
6	What's the latest news today?	5 Latest News about Universitas Internasional Batam: -EC's Prelude: Start Your Journey @EnglishClub_UIB (Date: Oct 7, 2021) Further news: https://www.uib.ac.id/en/ecs-prelude-start-your-journey-englishclub_uib-2/ ...	News	Success
7	What are offered scholarship in UIB?	There are 6 types of scholarships in UIB, which are: 1. Bright Scholarship 2. Independent Scholarship 3. Achievement Scholarship 1 4. Achievement Scholarship 2 5. Achievement Scholarship 3 6. Achievement Scholarship 4 Please visit Undergraduate Scholarship Programs for further information.	Scholar-ship	Success
8	Research institution and community service	Universitas Internasional Batam has some bureaus like: 1) Research Institutions and Community Service 2) Public Relations and Cooperation Bureau 3) UIB Center of Language and Culture 4) English Center 5) Mandarin Center 6) Bureau of Student and Alumni Development 7) Training & Certification Those Bureau will help connecting students to build necessary academic skills.	Bureau	Success
9	Can I register online to be a new student of UIB?	You can now register online as a new student at Universitas Internasional Batam. Just prepare your legal documents and go to Registration Site Or go to Online Registration Guide Page for complete guide.	Registra-tion	Success
10	I can't find topic or article on UIB Website, can you help?	Hi, E-service is developed to help you solve your problem or confusion at Universitas Internasional Batam. Please mention specific problem that you're facing. We'll trying to help you out :) Visit E-service Site for more services.	E-service	Success

Discussion

Black Box Testing is implemented for testing the chatbot. Black Box Testing works by focusing on the output result, in other words, it ignores the internal system's mechanism. This technique does not care about how the system is working to process the input, it's literally just to determine whether an error is encountered or any fail is occurred as of the result is not as expected (Ependi & Oktaviani, 2017; Firdaus et al., 2020; Supriyono, 2020). The user and the chatbot will have text to text communication as seen in Table 6. It shows the user's question example with its actual answers from the chatbot. It's also completed with which intent the response is classified and testing status whether it is success or fail (Fryer et al., 2019; Patel et al., 2019; Pi & Majid, 2020). Status will be stated as success if the answer meets the expected responses in accordance with template and response prepared at data pre-processing stage (Chandra & Suyanto, 2019). Answers could be in the form of URL that leads to a specific page of UIB website with an eye to ease user getting further information about their question.

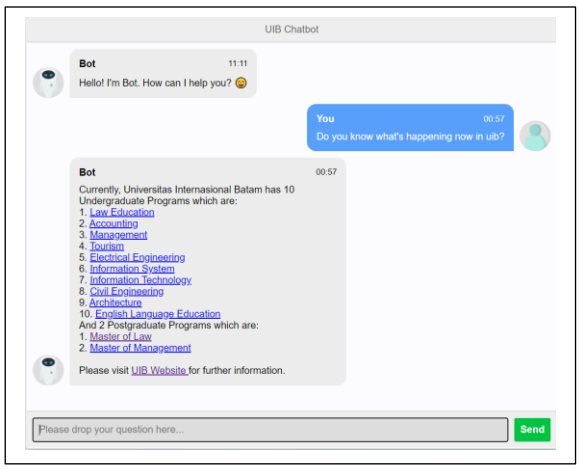

Figure 6. Failed Testing Result

One of the limitations of current research is that the chatbot only supports the English language. Improvement is needed in future research to support Indonesian language, like a previous research that implemented Chatbot with NLP approach using NLTK by (Patel et al., 2019; Prianto & Harani, 2021). Since UIB's website also supports those 2 languages, it can improve chatbot functionality. Based on the black box testing results, some questions are found not answered accurately as shown in Figure 6. It's happened due to term difference that's used by users and is not trained to the chatbot. The answer is supposed to be from 'news' or 'events' intents since 'happening' has a similar meaning with 'event' or 'occasion'. Varied word used in the collection of question patterns on each intent will prevent inaccurate answers due to misleading term (Arthana et al., 2021).

4. CONCLUSION

A communication system in the form of a web-based chatbot helps students and others to drop questions about Universitas Internasional Batam in colloquial language since it uses Natural Language Processing approach. It also helps users to obtain desired information in real-time, anytime and anywhere without having to visit colleges. This research presents the details about design and implementation of the successfully built chatbot. Through the testing process using 100 questions that are evenly distributed for each intent which some. This chatbot has the ability to answer questions with 86% accuracy rate.

5. REFERENCES

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the Communication Between Citizens and Government through AI-Guided Chatbots. *Government Information Quarterly*, *36*(2), 358–367. https://doi.org/10.1016/j.giq.2018.10.001.

Anggreani, D., Putri, D. P. I., Handayani, A. N., & Azis, H. (2020). Knuth Morris Pratt Algorithm in Enrekang-Indonesian Language Translator. *4th International Conference on Vocational Education and Training, ICOVET 2020*, 144–148. https://doi.org/10.1109/ICOVET50258.2020.9230139.

Anghelescu, P., & Nicolaescu, S. V. (2019). Chatbot Application using Search Engines and Teaching Methods. *Proceedings of the 10th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2018*, 1–6. https://doi.org/10.1109/ECAI.2018.8678948.

Arisandi, D., Indra, Z., & Kartini, K. (2021). Mengidentifikasi Hoax pada Hasil Pencarian Berita Online dengan Teknik Web Scraping dan Algoritma C4.5. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, *6*(2), 130–137. https://doi.org/10.36341/rabit.v6i2.690.

Arthana, I. K. R., Joni, L., Dewi, E., Seputra, K. A., & Wayan, N. (2021). Undiksha Virtual Assistant (Shavira) : Integration Frequency Asked Question with Rasa Framework. *JST (Jurnal Sains Dan Teknologi)*, *10*(2), 264–273. https://doi.org/10.23887/jst-undiksha.v10i2.39863.

Bertram, C., Weiss, Z., Zachrich, L., & Ziai, R. (2021). Artificial Intelligence in History Education Linguistic Content and Complexity Analyses of Student Writings in The CAHisT Project (Computational Assessment of Historical Thinking). *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2021.100038.

Chandra, Y. W., & Suyanto, S. (2019). Indonesian Chatbot of University Admission Using a Question Answering System Based on Sequence-to-Sequence Model. *Procedia Computer Science*, *157*, 367–374. https://doi.org/10.1016/j.procs.2019.08.179.

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the Impact of

Anthropomorphism on Consumer Response to Chatbots from The Perspective of Trust and Relationship Norms. *Information Processing & Management*, *59*(3). https://doi.org/10.1016/j.ipm.2022.102940.

Dahiya, M. (2017). A Tool of Conversation: Chatbot, International Journal of Computer Sciences and Engineering. *International Journal of Computer Sciences and Engineering (JCSE)*, *5*(5).

Danniels, E., Pyle, A., & DeLuca, C. (2020). The Role of Technology in Supporting Classroom Assessment in Play-Based Kindergarten. *Teaching and Teacher Education*, *88*. https://doi.org/10.1016/j.tate.2019.102966.

Dawood, S. S., & Barakat, S. A. (2020). Empirical Performance Evaluation of Knuth Morris Pratt and Boyer Moore String Matching Algorithms. *Journal of University of Duhok*, *32*(1), 134–143. https://doi.org/10.26682/sjuod.2020.23.1.14.

Ependi, U., & Oktaviani, N. (2017). Abstract Keyword Searching with Knuth Morris Pratt Algorithm. *Scientific Journal of Informatics*, *4*(2), 150–157. https://doi.org/10.15294/sji.v4i2.9797.

Eze, S. C., Chinedu-Eze, V. C., & Bello, A. O. (2018). The Utilisation of E-Learning Facilities in The Educational Delivery system of Nigeria: a study of M-University. *International Journal of Educational Technology in Higher Education*, *15*(1). https://doi.org/10.1186/s41239-018-0116-z.

Firdaus, V. A. H., Saputra, P. Y., & Suprianto, D. (2020). Intelligence Chatbot for Indonesian Law on Electronic Information and Transaction. *IOP Conference Series: Materials Science and Engineering*, *830*(2). https://doi.org/10.1088/1757-899X/830/2/022089.

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot Learning Partners: Connecting Learning Experiences, Interest and Competence. *Computers in Human Behavior*, *93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023.

Grandi, R., Falconi, R., & Melchiorri, C. (2011). UniBot Remote Laboratory: A Scalable Web-Based Set-up for Education and Experimental Activities in Robotics. *IFAC Proceedings*, *44*(1). https://doi.org/10.3182/20110828-6-IT-1002.03103.

Hajikhani, A., Pukelis, L., Suominen, A., Ashouri, S., Schubert, T., Notten, A., & Cunningham, S. W. (2022). Connecting Firm's Web Scraped Textual Content to Body of Science: Utilizing Microsoft Academic Graph Hierarchical Topic Modeling. *MethodsX*, *9*. https://doi.org/10.1016/j.mex.2022.101650.

Jabeen, H. (2019). *Making Web Crawlers Using Scrapy for Python*. DataCamp.

Kakani, V., Nguyen, V. H., Kumar, B. P., Kim, H., & Pasupuleti, V. R. (2020). A Critical Review on Computer Vision and Artificial Intelligence in Food Industry. In *Journal of Agriculture and Food Research* (Vol. 2). https://doi.org/10.1016/j.jafr.2020.100033.

Kalnoor, G., & Agarkhed, J. (2018). Detection of Intruder using KMP Pattern Matching Technique in Wireless Sensor Networks. *Procedia Computer Science*, *125*. https://doi.org/10.1016/j.procs.2017.12.026.

Khader, M., Awajan, A., & Al-Naymat, G. (2018). The Effects of Natural Language Processing on Big Data Analysis: Sentiment Analysis Case Study. *ACIT 2018 - 19th International Arab Conference on Information Technology*, 1–7. https://doi.org/10.1109/ACIT.2018.8672697.

Manuhutu, & Wattimena. (2019). Perancangan Sistem Informasi Konsultasi Akademik Berbasis Website. *Jurnal Sistem Informasi Bisnis*, *9*(2). https://doi.org/10.21456/vol9iss2pp149-156.

Mpungose, C. B. (2021). Lecturers' Reflections on Use of Zoom Vdeo Conferencing Technology for E-Learning at a South African University in The Context of Coronavirus. *African Identities*. https://doi.org/10.1080/14725843.2021.1902268.

Muehlethaler, C., & Albert, R. (2021). Collecting Data on Textiles from The Internet Using Web Crawling and Web Scraping Tools. *Forensic Science International*, *322*. https://doi.org/10.1016/j.forsciint.2021.110753.

Patel, N. P., Parikh, D. R., Patel, D. A., & Patel, R. R. (2019). AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT). *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019*, 148–150. https://doi.org/10.1109/ICECA.2019.8822176.

Pi, S. N. M. S., & Majid, M. A. (2020). Components of Smart Chatbot Academic Model for a University Website Siti. *Emerging Technology in Computing, Communication and Electronics (ETCCE)*, 1–6. https://doi.org/10.1109/ETCCE51779.2020.9350903.

Prianto, C., & Harani, N. H. (2021). The Covid-19 Chatbot Application Using A Natural Language Processing Approach. *IJISTECH*, *5*(36), 198–206. https://doi.org/10.30645/ijistech.v5i2.133.

Rahim, R., Zulkarnain, I., & Jaya, H. (2017). A review: Search Visualization with Knuth Morris Pratt Algorithm. *IOP Conference Series: Materials Science and Engineering*, *237*(1). https://doi.org/10.1088/1757-899X/237/1/012026.

Rasmila. (2018). Evaluasi Website dengan Menggunakan System Usability Scale (SUS) pada Perguruan

Tinggi Swasta di Palembang. *JUSIFO (Jurnal Sistem Informasi)*, *4*(1).

Sharma, A. K., Bajpai, B., Adhvaryu, R., Pankajkumar, S. D., & Gordhanbhai, P. P. (2021). An Efficient Approach of Product Recommendation System using NLP Technique. *Materials Today: Proceedings*, *6*. https://doi.org/10.1016/j.matpr.2021.07.371.

Sharma, M., Luthra, S., Joshi, S., & Kumar, A. (2021). Implementing Challenges of Artificial Intelligence: Evidence from Public Manufacturing Sector of An Emerging Economy. *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2021.101624.

Supriyono, S. (2020). Software Testing with The Approach of Blackbox Testing on the Academic Information System. *International Journal of Information System and Technology*, *3*(2), 227–233. https://doi.org/10.30645/ijistech.v3i2.54.

Surabhi, M. C. (2013). Natural Language Processing Future. *International Conference on Optical Imaging Sensor and Security (ICOSS)*, 1–3. https://doi.org/10.1109/ICOISS.2013.6678407.

Tovar Viera, R., Velasco Sánchez, Diego IsmaelTovar Viera, R., & Velasco Sánchez, D. I. (2020). Research on Technology Competencies in EFL Language Instructors: Technology-Pedagogy-Content in Language Teaching. *Script Journal: Journal of Linguistics and English Teaching*, *5*(1), 32–43. https://doi.org/10.24903/sj.v5i1.414.

Wahyuni, S., Reswita, & Afidah, M. (2020). Pengembangan Model Pembelajaran Sains , Technology , Art , Engineering and Mathemathic pada Kurikulum PAUD. *Jurnal Golden Age, Universitas Hamzanwadi*, *04*(2), 297–309. https://doi.org/10.29408/goldenage.v4i02.2441.

Warnars, H. L. H. S., Aurellia, J., & Saputra, K. (2021). Translation Learning Tool for Local Language to Bahasa Indonesia using Knuth-Morris-Pratt Algorithm. *TEM Journal*, *10*(1), 55–62. https://doi.org/10.18421/TEM101-07.

Wijaya, H. D., Gunawan, W., Avrizal, R., & Arif, S. M. (2020). Designing Chatbot for College Information Management. *IJISCS (International Journal of Information System and Computer Science)*, *4*(1), 8–13.

Wilujeng, I., Tadeko, N., & Dwandaru, W. S. B. (2020). Website-Based Technological Pedagogical and Content Knowledge for Learning Preparation of Science Teachers. *Cakrawala Pendidikan*, *39*(3), 545–559. https://doi.org/10.21831/cp.v39i3.31228.