Designing A Chatbot Application For Web-Based English Learning Using Boyer Moore Algorithm

Darwis Robinson Manalu[1], Marlyna Infryanty Hutapea[2], Richardo Lumban Raja[3], Arina Prima Silalahi[4], Harlen Gilbert Simanullang[5], Erwin Panggabean6

[1,2,3,4,5]Department of Informatics Engineering, Faculty of Computer Science, Universitas Methodist Indonesia
[6]Department of Information Technology, STMIK Pelita Nusantara Medan, Indonesia

Article Info

ABSTRACT

In the current era, the ability to use English is very important because English is not only used to communicate but also as a source of information. However, many students face challenges during the English learning process, including monotonous and less interactive learning methods. Therefore, an interactive and accessible Chatbot-based application is needed to help learn English. In this application, Boyer-Moore algorithm is used for keyword pattern matching of questions with knowledge base. Boyer-Moore algorithm is a string search algorithm that optimizes character matching from right to left. The results showed that the application of Chatbot application using Boyer-Moore Algorithm can answer questions with an accuracy rate of about 96% from a total of 100 questions tested.

Corresponding Author:

Darwis Robinson Manalu,
Department of Informatics Engineering
Faculty of Computer Science
Universitas Methodist Indonesia,
Jl. Hang Tuah no 8 Madras Hulu Medan Polonia Medan, Propinsi Sumatera Utara, Indonesia.
Email: manaludarwis@gmail.com

1. INTRODUCTION

English is widely spoken in many countries and is used in many fields, including education, technology, business, industry, politics, trade and others. In the current era, skills in using English are needed. Having skills in using English can open up opportunities to work on a multinational scale, participate in the global community, and read international research. Based on research conducted by Education First (EF) with the topic of discussion is the English Language Proficiency Index in 2022, Indonesia is ranked 81 (low) out of 111 countries. In 2021, Indonesia is ranked 80th and in 2020 it is ranked 74th. Based on this data, it can be concluded that the ability of Indonesian people to use English is still lagging behind other countries. However, in the process of learning English, there are various problems faced by students. One of the problems that often arises is the monotonous or less interactive teaching method, which comes from the teacher [1,2]. Based on this problem, it is important to develop an application that can help learning become more interesting to learn and easily accessible at any time in learning English. One solution is to utilize Artificial Intelligence and Chatbot [3 ,4,5]

Artificial Intelligence is a part of computer science that aims to develop computer systems to have abilities and behaviors similar to humans [4,5,6,7,8,9,10].

One of the fields of artificial intelligence can be implemented in a chatbot application. Chatbot is a program designed to allow humans and computers to communicate with each other using natural language just as humans communicate with other humans [13]. One application of the Boyer-Moore Algorithm can be applied to Chatbot. Where Boyer Moore will perform pattern matching on question keywords, then match them with keywords contained in the knowledge base. Boyer-Moore is a string search algorithm that starts the process of matching characters from the right and then shifts to the left[14].

Chatbot is a series of programs that allow humans and computer systems to interact just as humans communicate with other humans [10, 11]. The term "chatbot" comes from "chatterbot", which is a term originally proposed by Michael Mauldin in 1997 to describe a robot that can be used by humans to chat. This technology is also known by other names such as dialog system, conversational agent, conversational interface, virtual assistant, and personal assistant [17]. The use of chatbots has been widely used in various fields, such as customer service in e-commerce, financial dialogue systems, virtual counseling services, and conversational agents in health consultations [18,19].

2.   RESEARCH METHOD

The Research and Development method is the method used in this research. This method is divided into several research steps, as for the steps in question can be explained in Figure 1 below :
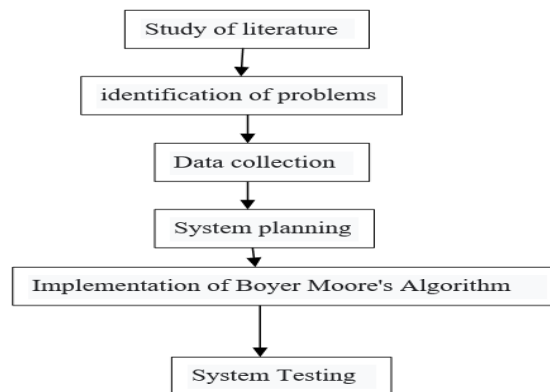

Fgure 1. Research Framework

Figure 1 explains that there are several steps to be taken. Below are the steps that will be carried out, namely:

1. Literature Study, The first thing to do is to collect information related to books related to the discussion of tenses in English.
2. Problem Identification, The problem that researchers identify is how to design applications that can help learn English, especially learning about tenses.
3. Data Collection, The process carried out at this stage is to conduct a literature study in order to find questions related to tenses and the correct answers to these questions.
4. System Design, In designing the system in question, the javascript programming language was used as well as the use of the MySQL database.
5. Boyer Moore Implementation, The next step is to apply the Boyer Moore algorithm to the system that has been designed. By applying Boyer Moore's algorithm, the chatbot application will become

more reliable and effective in understanding and responding to user input, providing relevant solutions or information, and improving the overall interaction between users and chatbots.

6. System Testing, Confusion Matrix was used in testing. The goal is to evaluate the performance of the algorithm used in the chatbot. The calculation used in determining the Confusion Matrix value is to use the following equation.

$$AC = \frac{TP + TN}{TP + FP + TN + FN}$$

Where acccuracy is the distance between the number of correctly identified questions and the total available questions.

2.1 Data Collection

The data collection process is taken from the Brainly website which is used as a reference to collect a list of questions about tenses in English and uses a book by M. Furqon& Desi Sugiarti with the title English skill booster: grammar, tenses, vocabulary, conversation as a reference to create answers to the questions made. Brainly is an online learning platform that facilitates users to openly ask and answer questions from other users related to school subjects. On the brainly website there are more than 4000 questions related to tenses, from these questions the author collects questions that are often asked by brainly users to be used as brain files. Table 1 is a sample of data from the brain file used in this study.

Table 1 Brain File Data

No.	Keywords.	Answer
1	What is *simple present tense*	It is a description of time used when describing activities or situations that recur regularly in both the present and the future.
2	*Simple present tense* formula	Subject + verb (infinitive form) + object is the formula used to form sentences.
3	*Simple present tense* example sentences	a) *She studies every day.* (She studies every day.) b) *They live in New York.* (They live in *New York.*) c) *They watch movies together on weekends.* (They watch movies together on weekends.)
4	What is *present continuous tense*	It is a *tense* that shows the action or state that is taking place at the time of speaking. This *tense* is also known as the *present progressive tense.*
5	Example of a *present continuous* tense sentence	a) *I* am *eating breakfast right now.* b) *They are watching a movie together.* (They are watching a movie together.) c) *We are meeting at the park tomorrow* morning. (We are meeting at the park tomorrow morning.)
6	*Present continuous tense* formula	By combining Subjects, auxiliary verbs and more.
7	What is *present perfect tense*	It is a tense that reflects past actions or events and still has an impact.
8	*Present* perfect *tense* formula	Subject + has/has + Verb 3 (in Past Participle Form).
9	Example of a *present perfect tense* sentence	a) *I have visited Bali three times.* (I have visited Bali three times.) b) *They have never tried sushi before.* (They have never tried sushi before.) c) *I have lived in Medan for 20 years.* (I have lived in Medan for 0 years.)
10	What is *present perfect continuous tense*	It is one of the *tenses* in English that refers to an action or event that was in the past and lasted until now, and its effects are still felt in the present.

In order for the system built to work well, it is necessary to process the data before the data is used. The system receives input in the form of questions, after receiving input, the input will go through a text preprocessing process, so that the text search process is easier to do. The text preprocessing process is carried out through three stages as described in Figure 2 below
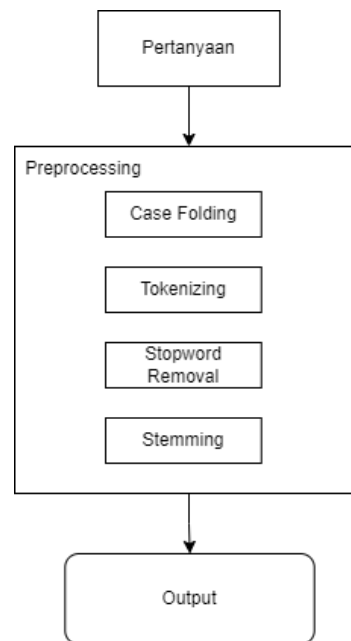

Fgure 2. Data Processing

2.2 Boyer Moore Algorithm

Boyer Moore algorithm is one of the most efficient string matching algorithms available as it actually finds matches in a sub-linear search time. This algorithm achieves this by simply scanning the key string from left to right. In case of a miss, the key string will be shifted a pre-calculated number of characters to the right until a match is made with the current character. Then, the next character that has not matched will be considered. Since the length of the key string and the position of the current character are known, the number of characters where a key string match can occur can be calculated [20]. Pseudocode of Boyer Moore Algorithm as follows :

```
procedure BoyerMooreSearch(
    input text: array[0..m-1] of char,
    input pattern: array[0..n-1] of char,
    input m, n: integer
) returns integer
Deklarasi:
badCharacterTable: array[0..255] of integer
goodSuffixTable: array[0..n-1] of integer
        j, shift, bmBcShift, bmGsShift: integer

makeBadCharacterTable(pattern, n, badCharacterTable)
makeGoodSuffixTable(pattern, n, goodSuffixTable)
    j := n - 1

    while j < m do
        if pattern[n - 1] = text[j] then
            if patternMatches(text, pattern, j - n + 1, n) then
                return (j - n + 1)
```

```
            else
bmBcShift := badCharacterTable[chartoint(text[j])] - n + j + 1
bmGsShift := goodSuffixTable[n - 1]
                shift := max(bmBcShift, bmGsShift)
                j := j + shift
            end if
        else
bmBcShift := badCharacterTable[chartoint(text[j])] - n + j + 1
bmGsShift := goodSuffixTable[n - 1]
            shift := max(bmBcShift, bmGsShift)
            j := j + shift
        end if
    end while

    return -1
end procedure
```

3. RESULTS AND DISCUSSIONS

The system consists of 2 parts, namely the admin and user parts. Admin consists of dashboard page, manage user data page, question data page, and feedback page. User consists of login page, register page and main page. Boyer Moore algorithm is used on the user's main page, when the user inputs a question in the conversation dialog, the question will be processed and a word match is made between the user's question and the brain file to the appropriate answer.

3.1. System Implementation
Login Page
Figure 3 is the user login user interface. Users who have never entered the website will be directed to the login page. The email and password used must be registered in the database in order to enter the main page of the website.
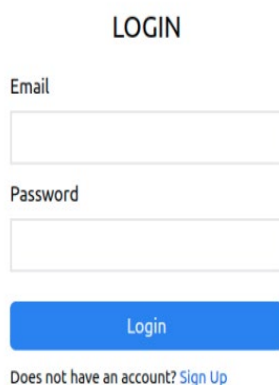

Figure 3. Login Page

Register Page
Figure 4 is the registration page. The register page is a page for users to register themselves so that they can log in and enter the website. On the register page, new users fill in username, email and password data.

Figure 4. Register Page

Home Page

Figure 5 is the user's main page where users can ask questions to the bot. On this page a conversation dialog will appear between the user and the bot and every question and answer from the bot will be saved as a conversation history
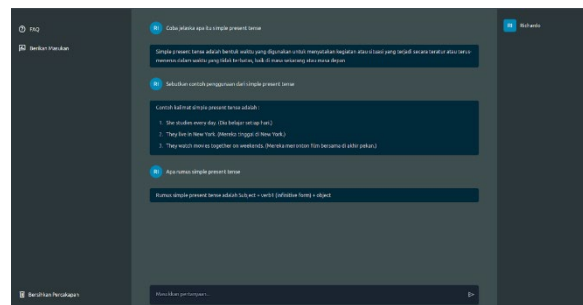
.

Figure 5. Home Page

3.2. Discussion

Testing the accuracy of the chatbot in providing answers that match the questions given by the user using Confusion Matrix testing. In testing, the test data used is 100 data. Based on the test results, the following results are obtained:

TP: Questions about tenses answered correctly by the chatbot

TN: Questions that are not about tenses and are answered incorrectly by the chatbot

FN: A question about tenses that was answered by a chatbot but the result was wrong.

FP: Questions about tenses and cannot be answered by chatbot.

	Positive	Negative
Positive	82 TP	3 FP
Negative	0 FN	12 TN

Based on the True Positive, True Negative, False Positive and False Negative values in the table above, the accuracy value can be calculated as follows

accuracy=   (TP+TN)/(TP+TN+FP+FN)
=   (82+12)/(82+12+3+0)
=   0.96*100%
=   96%

Based on the above calculation, the application of Boyer Moore algorithm to perform string matching in English learning chatbot application has an accuracy of 96%. The accuracy result is also influenced by many things. For example, the length and complexity of the strings being matched. If there are many questions or phrases with complex and long structures in the chatbot application, the Boyer Moore algorithm may have difficulty in matching the right words or phrases quickly and accurately. In addition, handling ambiguous keywords and writing or spelling errors can also affect the accuracy of the algorithm. Therefore, good implementation, selection of an efficient indexing method, and the size and quality of the dataset used also play an important role in improving the accuracy of the Boyer Moore algorithm in English learning chatbot applications.

In interpreting the accuracy results that researchers get, it is important to compare them with the accuracy results that have been achieved by similar research that has been done before. The comparison of the accuracy results of the application of the algorithms used can be seen in Table 3 below.

Table 2. Comparison of Accuracy Results

No.	Title	Author	Accuracy Result
1	Chatbot Application Using *Boyer Moore* Method (Case Study: C3 Ptipd Uin Sultan Syarif Kasim Riau)	(Angriyani 2019)	96%
2	Implementation of *Boyer Moore* Algorithm on Medicinal Plants *Chatbot*	(Ibrahim 2021)	99.41%.
3	Application of Boyer Moore Algorithm to Ustaz Abdul Somad Chatbot	(Ari Ismanto, 2021)	79%

Based on these three pieces of information, it can be concluded that the implementation of Boyer Moore algorithm on the chatbot shows various levels of accuracy, namely 79%, 96%, and 99.41%. The implementation on the Medicinal Plants chatbot achieved the highest level of accuracy, while the specific application for Ustaz Abdul Somad had a lower level of accuracy. However, the information provided is not enough to comprehensively evaluate the performance and effectiveness of each of these chatbot applications because the accuracy rate is only one factor that can be used to evaluate the performance of a chatbot. There are other aspects that need to be considered, such as response speed, ability to understand complex questions, and suitability to a particular context or domain.

4.   CONCLUSION

After planning, implementing, and testing the application, conclusions are drawn, namely:

1.   A chatbot application for English language learning has been successfully built and successfully answers questions related to 16 tenses using the boyermoore algorithm.
2.   Based on the results of confusion matrix testing, the chatbot accuracy rate reached 96%. The number of keywords provided in the brain file affects the accuracy of the chatbot,

so the more keywords provided, the more relevant the response provided by the chatbot.

REFERENCES

[1]   G. Tambunsaribu and Y. Galingging, "Masalah yang dihadapi pelajar bahasa Inggris dalam memahami pelajaran bahasa Inggris," *Dialekt. J. bahasa, sastra dan budaya*, vol. 8, no. 1, pp. 30–41, 2021.

[2]   D. Adhar, U. Indriani, E. Panggabean, A. Sabir, and H. Kurniawan, "Application of the Fisher-Yates Shuffle Algorithm in the Game Matching the World Monument Picture," in *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2020, pp. 1–6.

[3]   S. Wahyuni, O. S. Sitompul, E. B. Nababan, and P. Sihombing, "Social Network Analysis Text Mining on Networks Publication Citation," in *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, 2021, pp. 35–39.

[4]   T. Hirosawa, Y. Harada, M. Yokose, T. Sakamoto, R. Kawamura, and T. Shimizu, "Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: A pilot study," *Int. J. Environ. Res. Public Health*, vol. 20, no. 4, p. 3378, 2023.

[5]   S. Altay, A.-S. Hacquin, C. Chevallier, and H. Mercier, "Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions.," *J. Exp. Psychol. Appl.*, vol. 29, no. 1, p. 52, 2023.

[6]   S. Wahyuni and M. Marbun, "Implementation of Data Mining In Predicting the Study Period of Student Using the Naïve Bayes Algorithm Implementation of Data Mining In Predicting the Study Period of Student Using the Naïve Bayes Algorithm," in *IOP Confrence Series: materials Science and engineering*, 2020, pp. 4–11.

[7]   O. S. Sitompul and E. B. Nababan, "Biased support vector machine and weighted-smote in handling class imbalance problem," 2018.

[8]   O. S. Sitompul, E. B. Nababan, and Z. Alim, "Adaptive distributed grid-partition in generating fuzzy rules," in *2017 11th International Conference on Information & Communication Technology and System (ICTS)*, 2017, pp. 119–124.

[9]   N. Ratama and M. Munawaroh, "Perancangan Sistem Informasi Sosial Learning Untuk Mendukung Pembangunan Kota Tangerang Dalam Meningkatkan Smart City Berbasis Android," *Sains dan Teknol. Inf.*, vol. 5, no. 2, pp. 59–67, 2019.

[10]   A. Khaliq and S. N. Sari, "PEMANFAATAN KERANGKA KERJA INVESTIGASI FORENSIK JARINGAN UNTUK IDENTIFIKASI SERANGAN JARINGAN MENGGUNAKAN SISTEM DETEKSI INTRUSI (IDS)," *J. Nas. Teknol. Komput.*, vol. 2, no. 3, pp. 150–158, 2022.

[11]   S. Wahyuni, kana S. Saragih, and M. I. Perangin-angin, "Implemntasi Metode Decision Tree C4.5 Untuk Menganalisa Mahasiswa Dop Out," *ethos*, vol. 6, no. 1, pp. 42–51, 2018.

[12]   S. Wahyuni and F. Wadly, "Application Of Inventory And Service Transactions On Web-Based Cv Medan Teknik using the Agile Kanban Method," *Int. J. Comput. Sci. Math. Eng.*, vol. 2, no. 1, 2023.

[13]   T. Sekarwangi, K. E. Sartono, A. Mustadi, and A. Abdulah, "The Effectiveness of Problem Based Learning-Based Interactive Multimedia for Elementary School Students," *Int. J. Elem. Educ.*, vol. 5, no. 2, pp. 308–314, 2021.

[14]   D. Surianto, D. Triyanto, and U. Ristian, "Penerapan Algoritma Boyer Moore dan Metode N-Gram Pada Aplikasi penyunting Naskah Teks Bahasa Indonesia Berbasis Web," *Coding J. Komput. dan Apl.*, vol. 8, no. 3, pp. 50–60, 2020.

[15]   P. Lee, S. Bubeck, and J. Petro, "Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine," *N. Engl. J. Med.*, vol. 388, no. 13, pp. 1233–1239, 2023.

[16]   L. Gkinko and A. Elbanna, "The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users," *Int. J. Inf. Manage.*, vol. 69, p. 102568, 2023.

[17]   S. Morsi, "Artificial Intelligence in Electronic Commerce: Investigating the Customers' Acceptance of Using Chatbots," *J. Syst. Manag. Sci.*, vol. 13, no. 3, pp. 156–176, 2023.

[18]   B. Luo, R. Y. K. Lau, C. Li, and Y. Si, "A critical review of state-of-the-art chatbot designs and applications," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 12, no. 1, p. e1434, 2022.

[19]   Husain, M. Zarlis, Z. Nasution, H. T. Sihotang, and S. Wahyuni, "Filsafat Ilmu Komputer Dan Cloud Computing Secara Etimologis," *J. Mantik Penusa*, vol. 2, no. 2, pp. 15–21, 2018.

[20]   A. Ojugo and A. O. Eboka, "Signature-based malware detection using approximate Boyer Moore string matching algorithm," *Int. J. Math. Sci. Comput.*, vol. 5, no. 3, pp. 49–62, 2019.