STM JOURNALS

Journal of
Software Engineering Tools & Technology Trends

ISSN: 2394-7292
Volume 11, Issue 1, 2024
January—April
DOI (Journal): 10.37591/JoSETTT

https://journals.stmjournals.com/josettt

Review JoSETTT

Developing a Chatbot System Utilizing Artificial Intelligence and Natural Language Processing

Archika Jain, Vikram Khandelwal*, Sohan Lal Gupta

Abstract

Software applications commonly feature a user interface that falls into broad categories, namely graphical user interface (GUI), text-based UI, or a blend of both. This interface is predominantly employed in web-based and desktop applications. A chatbot, designed to engage with users, operates through the storage and retrieval of session data. It proves particularly beneficial in situations where obtaining information about individuals who are not affiliated as students or employees of a college can be challenging. By leveraging a chatbot, a website can effectively furnish valuable information to its users. A user-engaging chatbot functions by storing and retrieving session data, demonstrating significant utility in scenarios, where gathering information about individuals not associated as students or employees of a college, poses challenges. Utilizing a chatbot enables a website to efficiently provide valuable information to its users.

Keywords: Lemmatization, artificial intelligence, knowledge base, chatbot, NLP, Word-net

INTRODUCTION

In the present era, chatbots are ubiquitous, serving as the primary responders to user queries within their designated domains. They function as the go-to source for addressing user questions in specific areas of operation. A chatbot is an artificial intelligence entity capable of engaging in conversations with humans, either through text or spoken dialogue in the case of voice-based interactions. Primarily utilized for acquiring information, it can operate on cell phones and local PCs, though it is commonly accessed via the internet. Chatbots possess the ability to be compelling, captivating, and spellbinding. Functioning as conversational agents, they interact with users within specific domains or on particular topics. Typically, a chatbot responds when a user asks a question or introduces a new discussion topic. Described as software agents simulating human entities, chatbots leverage AI and NLP to answer user queries, drawing upon a predefined knowledge base to generate responses.

The universal inquiry system serves as a rapid, standardized, and informative widget designed to elevate the overall user experience of the website, providing users with accurate information. The bot systematically analyzes user queries, comprehends messages, and responds appropriately, employing AI and NLP technologies. This approach not only saves users' time and effort but also ensures they receive effective answers. The system maintains a log file to record inputs that the chatbot could not handle. The administrator can review this log and integrate responses to relevant sentences into the knowledge base. This contributes to the continual improvement of the chatbot's knowledge system, encompassing its databases.

***Author for Correspondence**
Vikram Khandelwal
E-mail: vikram.khandelwal@skit.ac.in

Assistant Professor, Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Jaipur, Rajasthan, India

Received Date: March 05, 2024
Accepted Date: April 02, 2024
Published Date: April 05, 2024

Citation: Archika Jain, Vikram Khandelwal, Sohan Lal Gupta. Developing a Chatbot System Utilizing Artificial Intelligence and Natural Language Processing. Journal of Software Engineering Tools & Technology Trends. 2024; 11(1): 24–29p.

The goals of this application include:

- Analyzing users' queries and comprehending their messages.
- Delivering highly effective answers to user queries.
- Saving users' time by eliminating the need to personally visit a general store for inquiries.
- Keeping users updated on general activities.
- Providing responses through an effective GUI, creating a conversational experience that simulates interaction with a real person.

LITERATURE REVIEW

Li *et al.* [1]

Concentrating specifically on evaluating the similarity between concise sentence-length texts, this approach introduces an algorithm that considers both semantic and word order information embedded in the sentences. The algorithm computes the semantic similarity of two sentences by leveraging data from a well-organized lexical database and incorporating corpus statistics. The utilization of a lexical database enables our approach to capture human common sense knowledge, while the integration of corpus statistics enhances its adaptability to diverse domains. This method is versatile and applicable to a range of tasks involving text knowledge representation and discovery [1].

Haller *et al.* [2]

They presented a method for pinpointing crucial details in texts depicting the life, including the personality, of a historical figure. This approach aims to create a conversational agent suitable for implementation in middle-school Computer-Supported Collaborative Learning (CSCL) scenarios [2].

Slave *et al.* [3]

A bot is educated and operates in accordance with predefined rules during training to respond to questions, constituting a rule-based approach. The creation of such bots is relatively uncomplicated using this rule-based methodology. However, limitations arise when the bot encounters questions whose patterns do not align with the established rules. The language utilized for crafting these bots is the Artificial Intelligence Markup Language (AIML), an XML-based language that empowers developers to articulate the rules guiding the bot's responses [3].

Ahmed *et al.* [4]

They demonstrated the practical implementation of a chatterbot capable of being applied across diverse domains such as education, healthcare, and route assistance [4].

The model employed is statistical, built upon the structure of AIML (Artificial Intelligent Markup Language) for training, and incorporates Microsoft voice synthesizer for speech recognition and natural language processing functionalities [5].

PROPOSED SYSTEM

Context Identification

Input text undergoes pre-processing to conform to the system's requirements, ensuring standardization. The identification of appropriate context is facilitated by analyzing the keywords used in the text.

Personal Query Response System (Module-1)

The user's authenticity is verified using their user ID and password when they received personal queries, such as CGPA or attendance. An appropriate response is generated if the user details are invalid. Upon successful authentication, the input text undergoes processing to extract relevant keywords. Utilizing these keywords, the system comprehends the user's information requirements and retrieves the necessary data from the database.

Response System (Module-2)

When a user engages informally with the bot, their input undergoes comparison with relevant patterns stored in AIML files. If a relevant response is identified, it is relayed to the user. Simultaneously, user information such as username and gender is retained. In instances where no matching pattern is discovered in AIML files, a random response is dispatched, signaling "Invalid Input".

System for Analysing Queries and Generating Responses (Module-3)

When a user seeks information about general topics, this module is responsible for delivering the response. Upon alignment of the input with a pattern in the AIML files, the corresponding response is activated. If there is no specific entry in the AIML files for the given query pattern, keywords are extracted from the input [6].

A sentence similarity algorithm (NLP) is then applied to the modified input, comparing it with questions from a predefined set for which answers are available. If a sentence is identified with confidence greater than 0.5, the corresponding question's answer is returned as the response. If no questions match the user input, the input is logged for system improvement by the administrator. The administrator has the option to integrate the answer to that query into the knowledge base. Meanwhile, a random response is conveyed to the user, indicating "Answer not available".

Context Reset

Once the user is satisfied with the bot's response and chooses not to proceed with the conversation, there is an option to either log out of the system or exit directly. Upon the user's exit from the system, all input parameters are automatically reset [7].

DESIGN AND IMPLEMENTATION

AIML

In developing our knowledge base for regular conversations, we employed AIML files to store pairs of questions and answers. In the course of engaging with our chatbot, user input is matched against patterns stored in AIML files, and the corresponding response is delivered [8]. The structure of a sample AIML file is as follows:

```
<aiml version = "1.0.1" encoding = "UTF-8"?>
<category>
   <pattern> HELLO USERNAME</pattern>
   <template> Hello User!</template>
</category>
```

Utilizing Word-Net for POS Tagging and Lemmatization

The extraction of information from the input text involved identifying keywords. For instance, in the phrase "What is the current placement scenario?" the keywords are "current", "placement", and "scenario". To group various inflected forms of words together, appropriate lemmas of the keywords were determined through lemmatization and POS tagging. For example, the forms "requiring", "require", and "required" should all map to the lemma "require". This task utilized the WordNet feature from the "click" package in Python [9].

Log File

A log file is retained to store inputs that the chatbot could not address. The administrator has the ability to review the log and incorporate responses to pertinent sentences into the knowledge base. This process contributes to the enhancement of the chatbot's knowledge system, including its databases.

Use Case Diagram: Figure 1 shows the role of user and admin by using use case diagram [10, 11].

Data Flow Diagram: Figure 2 demonstrates the data flow diagram for the user and admin.

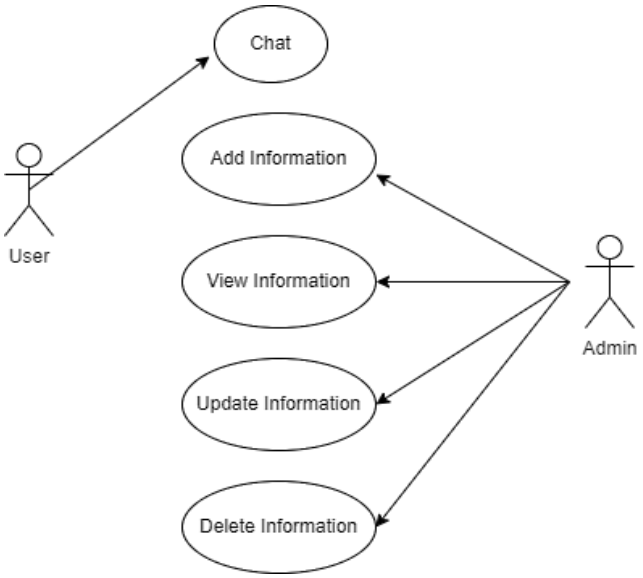

Figure 1. User and admin roles in use case diagram.

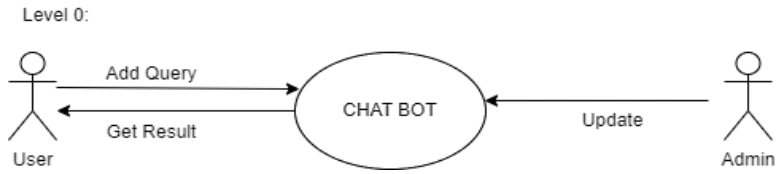

Figure 2. Data flow diagram of user and admin.

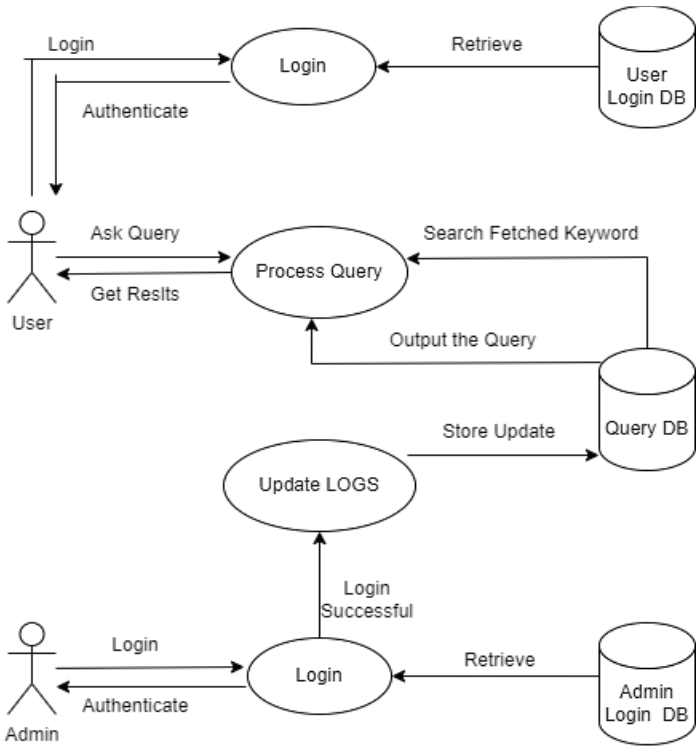

Figure 3. Chatbot system First level DFD.

Figure 3 demonstrates the chatbot system for the first level DFD.

Use Case Diagram: Figure 4 shows use case diagram of context identification.

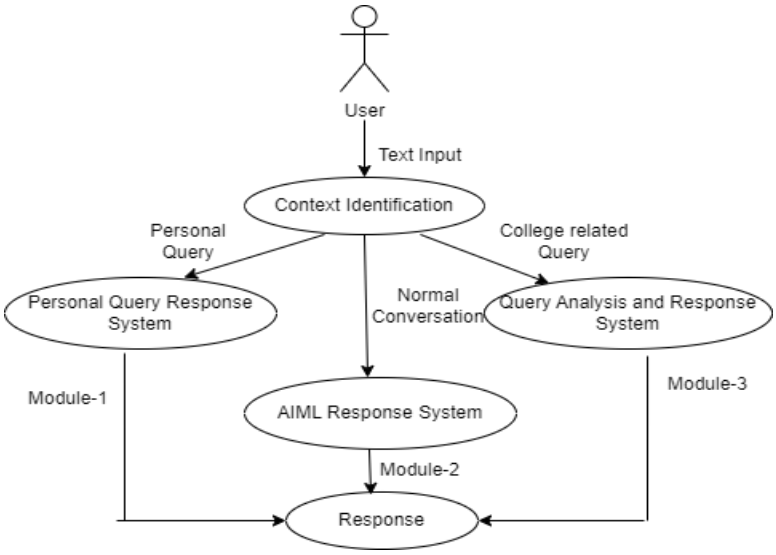

Figure 4. Context identification use case diagram.

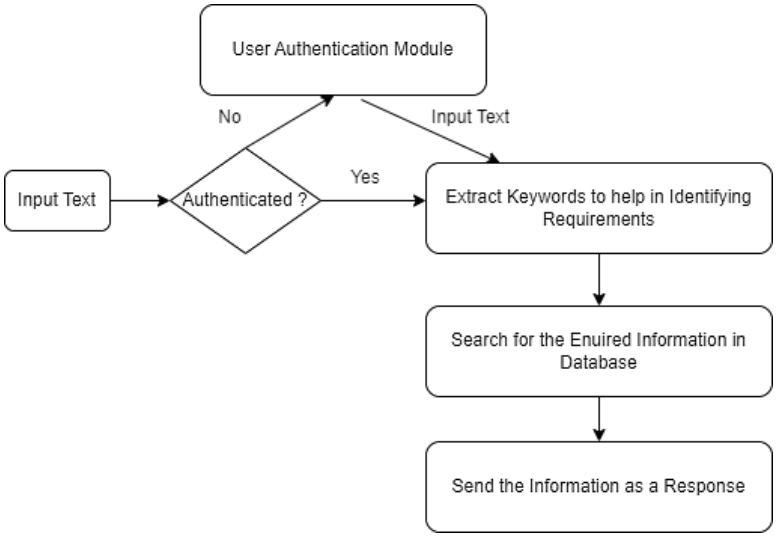

Figure 5. Personal query response activity diagram.

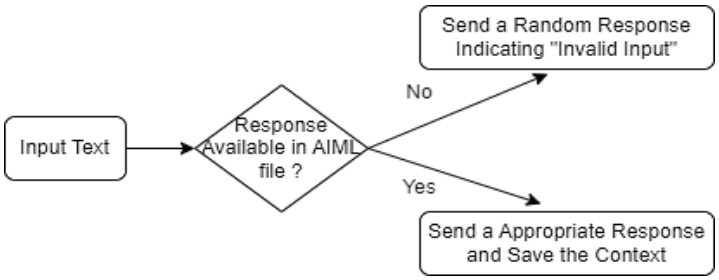

Figure 6. Normal conversation response activity diagram.

Activity Diagram: Figure 5 demonstrates the personal query response activity diagram of Module-1.

Figure 6 shows the normal conversation response activity diagram of Module-2.

CONCLUSION AND FUTURE WORK

Gaining access to extensive data often poses a challenge, requiring navigation through various forms and windows. The overarching goal of the chatbot is to streamline this complexity, offering a unified

and user-friendly interface that serves the needs of the general users. Chatbot system gives the output in human understanding form. Its architecture uses natural language for giving the output communication by combining the language model and computational algorithm. Both general users and employees are free to submit their queries. The chatbot quickly and effectively searches for answers to these queries, retrieving pertinent links. The background research involves analyzing the conversation process to identify relevant keywords associated with the query. Furthermore, a two-part interface has been developed, one for users and another for the administrator. Rather than relying on an AIML-based bot, alternative algorithms can be employed. The incorporation of voice-based queries is another possibility, where users provide voice input and receive text output from the system. After successfully deploying the chatbot in a general domain, expansion into other domains such as medical, forensic, sports, etc., can be considered. This approach proves advantageous across various fields, enabling efficient access to relevant information without the need for extensive time investment or sorting processes.

REFERENCES

1. Yuhua Li, David McLean, Bandar Zuhair A, O'Shea James D, Keeley Crockett. Sentence Similarity Based on Semantic Nets and Corpus Statistics. IEEE Trans Knowl Data Eng. 2006 Aug; 18(8): 1138–1150.
2. Emanuela Haller, Traian Rebedea. Designing a Chat-bot that Simulates a Historical Figure. IEEE Conference Publications. 2013 Jul; 582–589.
3. Pratik Slave, Vishruta Patil, Vyankatesh Gaikwad, Girish Wadhwa. General Enquiry Chat Bot. Int J Recent Innov Trends Comput Commun. 2015 Mar; 5(3): 463–466.
4. Imran Ahmed, Shikha Singh. AIML Based Voice-Enabled Artificial Intelligent Chatterbot. International Journal of u- and e-Service, Science, and Technology (IJUESST). 2015; 8(2): 375–384.
5. Kulkarni Chaitali S, Bhavsar Amruta U, Pingale Savita R, Kumbhar Satish S. BANK CHATBOT - An Intelligent Assistant System Using NLP and Machine Learning. Int Res J Eng Technol. 2017 May; 4(5): 2374–2377.
6. Studente S, Ellis S. Enhancing the Online Student Experience Through Creating Learning Communities—The Benefits of Chatbots in Higher Education. In: McKenzie S, Garivaldis F, Dyer KR, editors. Tertiary Online Teaching and Learning. Springer, Singapore; 2020. DOI: 10.1007/978-981-15-8928-7_3
7. Wikipedia. Verbot. https://en.wikipedia.org/wiki/Verbot
8. Kumar Shivam, Khan Saud, Manav Sharma, Saurav Vashishth, Sheetal Patil. Chatbot for College Website. Int J Comput Technol. 2018 Jun; 5(6): 74–77.
9. Ranoliya BR, Raghuwanshi N, Singh S. Chatbot for university-related FAQs. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi. 2017; 1525–1530. doi: 10.1109/ICACCI.2017.8126057.
10. Abdul-Kader Sameera A, John Woods. Question Answer System for Online Feedable New Born Chatbot. IEEE Intelligent Systems Conference, London, UK. 2017 Sep; 863–869.
11. Ming-Hsiang Su, Chung-Hsien Wu, Kun-Yi Huang, Qian-Bei Hong, Hsin-Min Wang. A Chatbot Using LSTM-based Multi- Layer Embedding for Elderly Care. IEEE International Conference on Orange Technologies (ICOT), Singapore. 2017; 70–74.

