Virtual Interface for College Enquiry Using AIML and Query Extraction Techniques and Algorithms

Sarang Digole[1], Harshal Date[2], Vipul Patil[3], Surbhi Bhaldand[4], Prof A.L. Golande[5]

Students, Department Computer Science[1,2,3,4]
Faculty, Department Computer Science[5]
JSPM's Rajarshi Shahu College of Engineering, Pune, Maharashtra, India
sarangdigole44@gmail.com
harshaldate2000@gmail.com
vipulpatil0411@gmail.com
surbhibhaldand1376@gmail.com
avinash.golande@gmail.com

Abstract

Students in the post-COVID world find it difficult to physically visit a college for any type of college information, such as scholarships, admissions procedures, and so on. It is indeed difficult to search and extract such information for someone who is not a student or employee there. Furthermore, the workload on office administration grows significantly, and it is no longer able to address more than 2-3 students' inquiries at a time. So, a VIRTUAL INTERFACE, such as a 'Chatbot' or a 'Voicebot' comes into the picture. Nowadays people to engage with systems in an escalating number of ways via chatbots and voice assistants rather than visiting or interacting manually. The new modes of communication would be chatbots, voice-bots, or voice assistants. This paper explains the advancement in Artificial Intelligence [AI], Machine Learning [ML], and Natural Language Processing [NLP] may assist us in developing a chatbot or voice-bot. The college inquiry system will respond by summarizing the query and then outputting replies, it will also supply the user with selective information. A college enquiry virtual system will give all responses important regarding various fields like examination cell, student admissions, academic & college notices, attendance of students, all placement related activity from placement cell and other miscellaneous fields.

INTRODUCTION:

In the present time and age, chat bots' technology is seen to be everywhere. Chatbots are a resource of solutions to consumers' inquiries in any industry where they operate. It is a computer software that mimics human communication by using AI which stands for Artificial Intelligence and NLP which is Natural Language Processing to interpret client requests and automate responses. The bot is useful to address questions about material that isn't immediately accessible on that website. It

helps consumers discover the information they need by replying to their inquiries and requests whether through text, speech, or both without requiring human participation. Chatbot technology is fairly widespread these days, from smart speakers at home to office social applications. "Virtual assistants" or "Virtual agents" are terms used to describe the latest AI chatbots. Most websites provide users with chat bots to guide them in navigating the website's features.

Chatbot fundamentals:

A chat bot is a machine with artificial intelligence that is capable of interacting with people. This might be a written dialogue or in the case of voice-based queries it can be a verbal one. They are mostly used to collect data. It is usually accessed over the internet. It has the potential to be compelling, interesting, and enthralling. It's a conversational agent that uses natural language phrases as input to speak with users in a certain sector or on a specific subject. A chat bot is often engaged when a user asks a question or initiates a new subject of conversation. Software agents that represent humans are known as chat bots. These are AI-enabled agents that can respond to specific questions using natural language processing (NLP). The construction of an answer to the inquiry is simplified by the usage of a predetermined knowledge base.

Chat bot for Institution:

For a variety of reasons, a college inquiry system is required including the slowness of college portals, the fact that an outsider would have no idea where to go for a certain piece of information, and the difficulty of obtaining data from outside the college's domain. The sensible solution to all of the system's flaws supports the system's demand. The college inquiry system will answer by analyzing the question and then responding; it will also provide selective information based on the user's choices. Admissions, assessments, notice boards, attendance, recruitment, and other associated sectors will all receive replies from a college system.

This project's objective is to build an intelligent chatbot system that will provide academic activities like as admissions, fees, scholarship information, departmental schedules, and information on required papers, among other things. Students will be able to get immediate responses to their questions in less time thanks to chat-bot technology.

RELATED WORK:

Recently, there has been a lot of research on this topic, and various researchers have published their findings.

The "Chatter Bot Algorithm" was devised by Michael Maudlin in 1994 and published in the book Julia, and it was used to answer the inquiries. The user must first log in to the Chat-Bot program, after which they will be able to submit their questions. When a user submits a question to the algorithm, the context of the inquiry gets identified, and natural language processing gets applied on it.

In 1995, ALICE was built by Richard Wallace, who was a pioneer in that field. At the start it was known as ALICEBOT. An AIML which stands for Artificial Intelligence Markup Language, which is an Extensible Markup Language [XML] schema, was used by the application which helps to establish dialogue rules. The application employs heuristic pattern matching to engage

on a conversation with living person via the Internet services. Alice was a lady who used to tell about her hobbies, interests, and other interesting details and activities she was able to do along with responding to people's questions.[1]

S. J. du Preez and M. Lall worked on a black box mechanism that enables all sorts of clients to interact with the server from any environment by managing the system of communication to and from the Web-service. They have explained a use of XML and AIML files in chatbot.[2]

The objective of a chatbot system, according to Tarun Lalwani and Shashank Bhalotia (2018), is to simulate digital communication among a person and a computer using natural human language, and to do so, its architecture combines a language model with a computational programme.[3]

Edward Zion Saji (2018) worked on a system, which specifically works on AI, that gathers information such as our browsing activity and interests and improves our internet browsing experience as well as our ability to do specific activities.[4]

Amir Vakili and Azadeh Shakery (2019) operated on a new system for use in retrieval-based chatbots. While forecasting the proportion of matching between a contender reply and a discussion context, the new architecture compares situations from the training set which are potential matches for the contender reply to the present context.[5]

Kyoko Sugisaki developed a web-based tool for text-based interactions that they created for research purposes in CMC which stands for 'Computer Mediated Communication'. It allows researchers to conduct linguistic anthropology using text-based real-time discussions. The system has two chat communication channels: 'quasi-synchron' and 'synchron modes,' as well as numerous typing indications. This system(tool) may also be utilized in the 'Wizard of Oz' research in HCI which is 'Human Computer Interaction' as well as for evaluating Virtual Assistants (Communication Systems) in NLP which is 'Natural Language Processing'.[6]

Punith.S. along with his colleagues (2020) have described about the software/tools used in making chatbot which explains use of DIALOGFLOW and NLP TOOLKIT in making of it. They have made use of different Machine Learning Algorithms like LSTM and Naive-Bayes.[7]

PROPOSED METHODOLOGY:

A. Login Authentication:
First of all, user authentication is required for the chatbot to work, so user have to login through their login credentials to access the chatbot. If user is visiting for first time, then it has to create a new account.

B. Context Generation:
Basically, pre-processing technique is applied to the user query where we remove irrelevant, redundant, and unnecessary data from a large amount of data [here user query] to convert it into useful and meaningful information as it is the first step in NLP for any model building.

There are various steps by which we can achieve the meaningful text which includes CLEANING to remove all punctuations, symbols, special characters, and stop words as well as to convert the whole text into either lowercase or uppercase since the data is composed of both uppercase and lowercase, TOKENISATION to convert all text into small tokens, STEMMING to

convert the different forms of a word into their base word and LEMMATIZATION to put together a word's inflected forms so that they'd be analyzed as a single item

C. Response Generation through AIML:

AIML files are mostly used under the Rule-based or Knowledge-based chatbots to simplify the modelling process of dialogue. When user asks any query, first of all it will be mapped with the AIML files stored in the system to find out whether the same pattern or query is available or not. If the patterns get matched it will give the corresponding response otherwise

If the user wants the answer related to something which is not included in AIML files then it won't capable to provide an answer and automatically that query will be send to the NLP model for further processing. Also, we have to manually update the knowledge of AIML files periodically.

D. Bot Response System:

As the query is not matched with any AIML files, it is passed to the NLP model where we use TFIDF algorithm which is nothing but "Term Frequency-Inverse Document Frequency" for finding out the cosine similarity of the input text. It finds out the similarity or confidence of the input query and based on that score appropriate response will be given from the corpus.

If question is not mapped with any information in the database, it will send random response like "Answer not available", etc. and the query will be sent to admin for review. Admin will review the query and add it in the corpus.

I] Login Authentication Diagram:
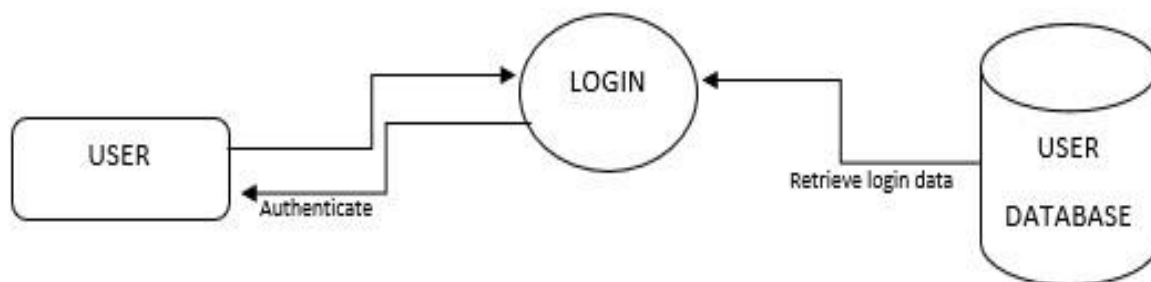
II] __General Work Flow Diagram:__
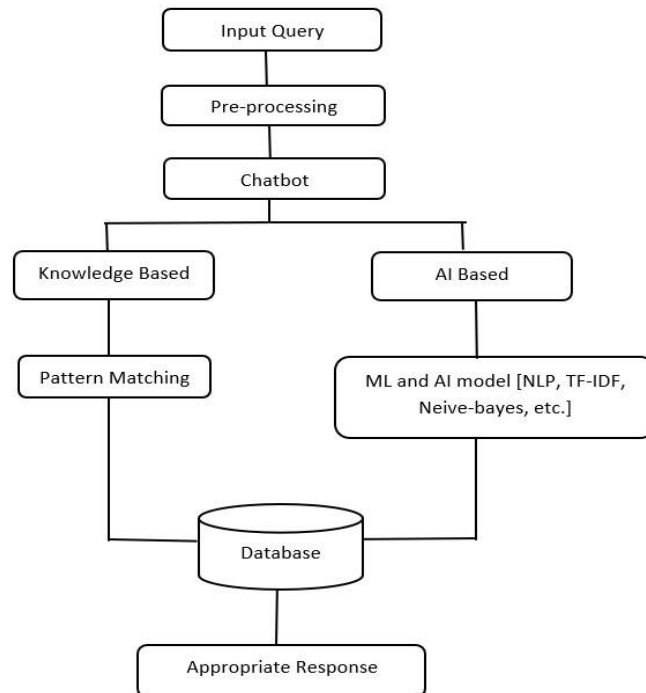
III] __Activity Diagram:__
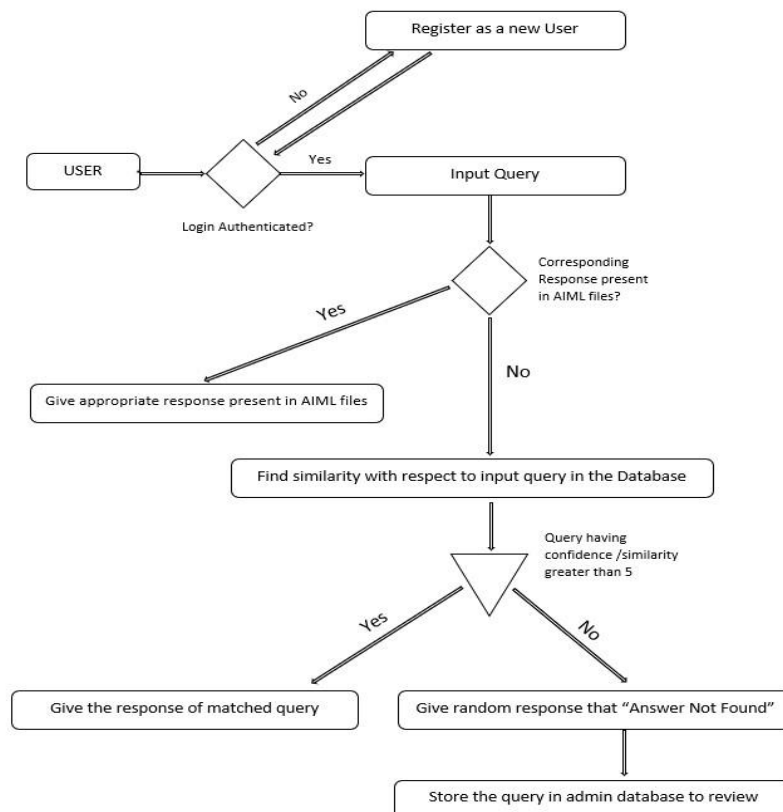
IMPLEMENTATION:

We have developed a Hybrid chatbot i.e., combination of rule-based and AI-based chatbot. To create our rule-based module we have used AIML and for AI-based module we have applied NLP techniques. When user converses with our chatbot, in first phase it will search pattern in AIML files, if found it will provide response otherwise query will be passed to second phase. In this phase, query will be processed and response will be generated by NLP techniques.

1] AIML:

To implement rule-based module of our hybrid chatbot we have used Artificial Intelligent Markup Language (AIML) files. This files store information in basic unit called 'Category'. Each category has an input query, an output solution, and a contest that may be entered. A rule-based module responds according to a connection between the query set by the user and response located in AIML file.

```
Example: -
<aiml version = "1.0.1" encoding = "UTF-8"?>
<category>
  <pattern> Founder of JSPM </pattern>
    <template> Dr. T. J. Sawant
    </template>
</category>
</aiml>
```

Result:

User: Founder of JSPM

Bot: Dr. T. J. Sawant

2] NLP:

Natural Language Processing technology is used to build AI-based module of our hybrid chatbot NLP technology aids in deciphering the purpose behind a user inquiry whose response isn't contained in an AIML file. The user query is pre-processed by different techniques like Tokenization, Stemming, Lemmatization, etc.

I] Tokenization:

In this step user query is broken down into small units called tokens. NLTK toolkit is used to perform this step. Below is the sample code for converting sentence into small tokens in python using NLTK library functions:

```
tokens_of_sentence = nltk.sent_tokenize(initial_text)

Tokens_of_words = nltk.word_tokenize(initial_text)
```

II] Lemmatization:

Information that we took out from the input query was performed by keyword extraction. For example, "What are different branches of JSPM?" contain "different", "branches" and "JSPM" as keywords. Lemmatization and POS tagging were used to categorize the various inflected forms of the words and select adequate lemmas. For example, plays, played and playing must match play. WordNet is the dictionary from the Python package "NLTK" was used for this step.

Below is the code snippet for lemmatization using NLTK library functions in python:

```
# Lematization :

lemmer = nltk.stem.WordNetLemmatizer()


def LemTokens(tokens):
    return [lemmer.lemmatize(token) for token in tokens]


def LemNormalize(text):
    return LemTokens(nltk.word_tokenize(text.lower().translate(remove_punct_dict)))
```

III] Semantic query similarity:

We have used TF-IDF algorithm to find the weights of words according to its importance in the document with respect to all other documents in our corpus. It's the product of two matrices, one is 'Term Frequency [TF]' which calculates mathematical figure of a term's frequency of occurrences among the database as well as document and the other is 'Inverse Document. Frequency [IDF]' that calculates how common or uncommon is the given word among all documents presents in the corpus. Rather than other models, this technique gives various values instead of only 0 or 1. So it becomes easy to find out which words are more important and less important.

For example, user sends query like "What is strength of JSPM?". So, the frequency of keywords like "strength", "JSPM", etc. will be calculated and based on maximum cosine-similarities found in database, it will select the response sentence.

1] TF (word, document) = No of times word appears in the document

/ No of times all the words appear in the document

2] IDF (word, documents) = log [total no of documents / total no of documents containing word]

3] TF-IDF (word) = TF (word, document) * IDF (word, documents)

APPLICATION:

1) It can be used as a part of college apps or websites so it saves time of teaching and non-teaching staffs.

2) Rule based module of our chatbot can be used to perform specific task like booking tickets, etc.

3) Enhanced AI-based chatbot will be used in most of the colleges and it can be used in various business-related industries.

4) It provides us with an easily accessible source of information without requiring any physical effort on our part.

RESULT AND OUTPUT:

Login Page:

Chatbot Interface:
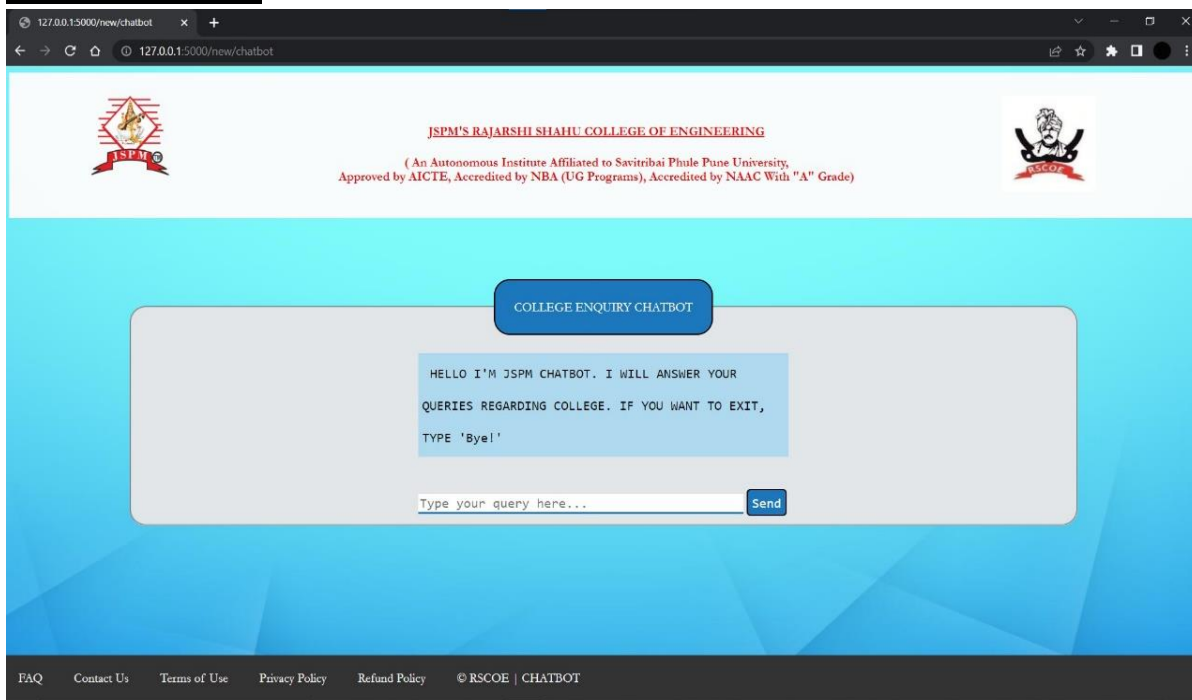
Login Authentication Database:
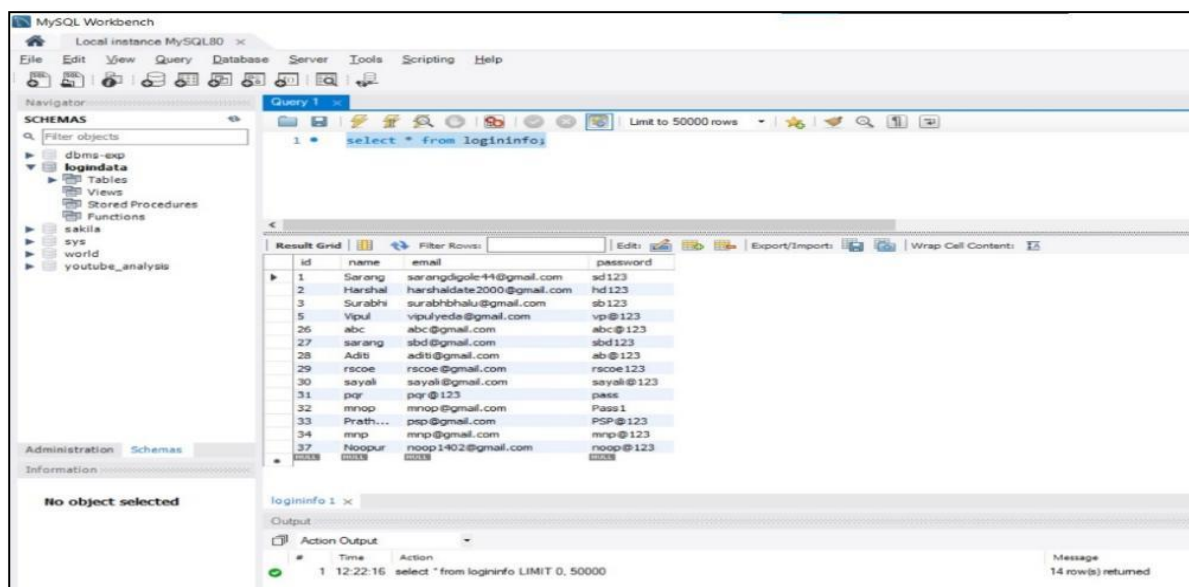
CONCLUSION:

The project's primary goal was to develop a web-based college enquiry chatbot to search and extract information as per user submitted queries. AIML files are used to store information related to college in the form of question-answer pairs. Various NLP algorithms had used to preprocess the queries and get the response. Students in college have complete freedom to type their questions.

The chat bot does a quick and efficient look for solutions to their questions and responds with the most relevant information.

REFERRENCES:

[1]   Atwell Eric and Bayan AbuShawar, "ALICE Chatbot: Trials and Outputs", Computación y Sistemas, No. 4, Volume 19, pp. 625–632, December 2015.

[2]   S. Sinha, Lall M., S.du Preez, "An intelligent web-based voice chat bot", in IEEE EUROCON 2009.

[3]   Bhalotia Shashank, Ashish Pal and Shreya B., "Implementation of a Chat Bot System using AI and NLP", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), Issue.3, Vol.6, May 2018.

[4]   Edward Zion Saji, "Simple Voice-Enabled chat-bot in Python", ICSCAN 2020.

[5]   Azadeh S. and Vakili Amir, Enriching Conversation Context in Retrieval-based Chatbots, in University of Tehran, Nov 2019.

[6]   Sugiski Kyoko, "Chat Bot Kit: A web-based tool to simulate text-based interactions between humans and with computers", Cornell University, arXiv:1911.00665,2 Nov 2019.

[7]   Punith S1, Chaitra B2, Veeranna K., "Research Paper on Chatbot for Student Admission Enquiry", JASET, Issue 1, Volume 3, September 2020.

[8]   Mrs. B.Hemalatha, "Interactive Chatbot Using AIML ", International Journal Of Advanced Networking & Applications, (IJANA).

[9]   Wibowo Ferry W., "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling" in International Conference on Intelligent Systems, Modelling and Simulation (ISMS), 2016 Jan.

[10]   Prof. Satish S. K., Amruta U. B. and Pingale Savita, "BANK CHAT BOT: An Intelligent Assistant System Using Machine Learning and NLP", International Research Journal of Engineering and Technology (IRJET), May 2017-Vol. 4th-Issue 5.

[11]   Mgr. Zemcik Tomas, "A Brief History of Chatbots", International Conference. on Artificial Intelligence, Control and Automation Engineering, AICAE 2019.

[12]   Patil Vedika, Gawade Harshala, "College Enquiry Chat-Bot System", IJERT 2020.

[13]   Harsha Pariyani, Anshika Sinha, Preeti Bhat, Roshni Rote, Prof. N. A. Mulla, "Literature Survey of Various Chatbots", JETIR 2020.

[14]   Mr. Gadhave Ganesh, Dnyaneshwar, Mr. Hambre Pankaj, Patil Rajnandini, Pawar Komal, "Web Based College Enquiry Chatbot", RJSE, issue 3, Volume 4, March 2020.