

INTELLIGENT CHATBOT FOR COLLEGE ENQUIRY SYSTEM : A SURVEY

Gandhar Khandagale¹, Meghana Wagh², Pranali Patil³, Hritvik Lad⁴, Prof. Satish Kuchiwale⁵

¹⁻⁴Students, Dept. of Computer Science, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India.

⁵Professor, Dept. of Computer Science, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India.

Abstract - College Enquiry ChatBot is an online service that allows users to converse with humans using artificial intelligence principles. The system deals with helping streamline the process of the admissions across various institutes. The system communicates with students in real time, removing any lingering reservations they may have about the admissions process. It is not realistic for institutes to set up a real-time question-answering assistant to assist admission candidates. So it will help students discover the institutes which they desire to go to. The goal of this project is to create an intelligent chat-bot system that will handle academic activities like admissions, fees, scholarship information, departmental timetables, and details on the documents that must be attached, among other things.

Key words : Pattern Matching A , Natural language processing (NLP), Naive Bayes Algorithm

I. INTRODUCTION

Chatbot is a computer programme that conducts an online chat conversation using text or text-to-speech technology. Inquiry to College ChatBot is an online application that converts with humans using artificial intelligence. "Eliza", "Cleverbot," and other comparable web applications have been produced in the past. This system is an application which responds to a student's query. The Student's Question is usually a conversation. The chat can be done in any format; there is no set format for conversing with the bot. For this we have used NLP in our project. The system gives users access to information about college-related events. The system has a user-friendly graphical interface. For query resolution, the system employs A.I. and M.L. The user can inquire about college-related activities, such as the date and time of annual day, sports day, and other cultural events. Save kids' and teachers' time, as well as extra personnel. Students can access all college documents, such as notices, study materials, and question papers, at any time and from any location, whether they are at college or not. In addition, employee labour will be reduced. It is effective communication between faculty and students.

The college chatbot's concept is based on Google Assistant, which will respond to student questions through a web application. It gives relevant links to the user's query. The user can get answers to their questions by clicking on the link. The chatbot uses a standard algorithm to extract data and display conversations between the human user and the bot. It answers questions about college activities such as results, timetables, significant notices, and placement notices, among other things. In most cases, the system finds the intents, or keywords.

II. MODULES

The procedures or strategies used to find, select, process, and analyse information about a topic are referred to as research methodology. The methodology portion of a research article allows the reader to critically examine the study's overall validity and dependability. Both qualitative and quantitative viewpoints are used in the proposed methodology. The system will have the following modules:

A. USER / ADMIN LOGIN

To access the various sections and ask questions or complain with the help of the bots, users must login with their credentials (i.e. User Id and Password) on the portal. During this process, users will be given the choice of selecting a language among Marathi, Hindi, or English. Admins (i.e. college authorities) will have access to all query logs through their own login. They will have the ability to edit or even add new answers to the system's inquiries.

B. INTERACTION WITH THE BOT

Instead of seeking assistance from the college helpdesk or college employees, users can communicate with the bot to ask or enquire about college-related events. This system is a programme that answers a question from a pupil. A dialogue frequently follows the Student's Question. There is no fixed format for chatting with the bot, so you can do it as you want. NLP was

employed in our project to do this. The user can query about college-related activities including annual day, sports day, and other cultural events, as well as the date and hour of each. Save time and money by reducing the number of children and teachers on the job. It is the ability of instructors and students to communicate effectively with one another.

C. TEXT TO SPEECH

This module explains how it turns text into speech, which aids in receiving a verbal answer. As a result, it is referred to as a verbal dialogue. In Python, there are numerous APIs for converting text to speech. The Google Text to Speech API, often known as the gTTS API, is one of these APIs. gTTS is a simple programme that translates entered text into audio that can be saved as an mp3 file. English, Hindi, Tamil, French, German, and many other languages are supported via the gTTS API.

III. SYSTEM ARCHITECTURE

The proposed System would be an application that provides answers to the queries provided by the scholar or the user. Users will simply ask questions using the chatbot that is currently in use. Students can chat in any manner they want; there is no requirement that the user adopt a specific format. The answers are applicable to what the user queries. If the answers are found to be invalid or not accessible, then those queries are held onto the unrequited table that is essentially created by the admin. Later those queries can be updated by the admin, just in case of urgency we are going to provide a message that “our representatives will get in touch with you shortly”.

To efficiently plan and develop the entire project, we used the spiral model in our project. The spiral approach is similar to incremental system development, except it focuses more on risk analysis. The spiral model aids in the effective planning and construction of the entire project. Planning, Design, Construct, and Evaluation are the four phases of the spiral model. Depending on the project's complexity and dangers, the project manager might adjust the number of stages required to complete the project.

The system flowchart of our project is shown in Fig.1 below.

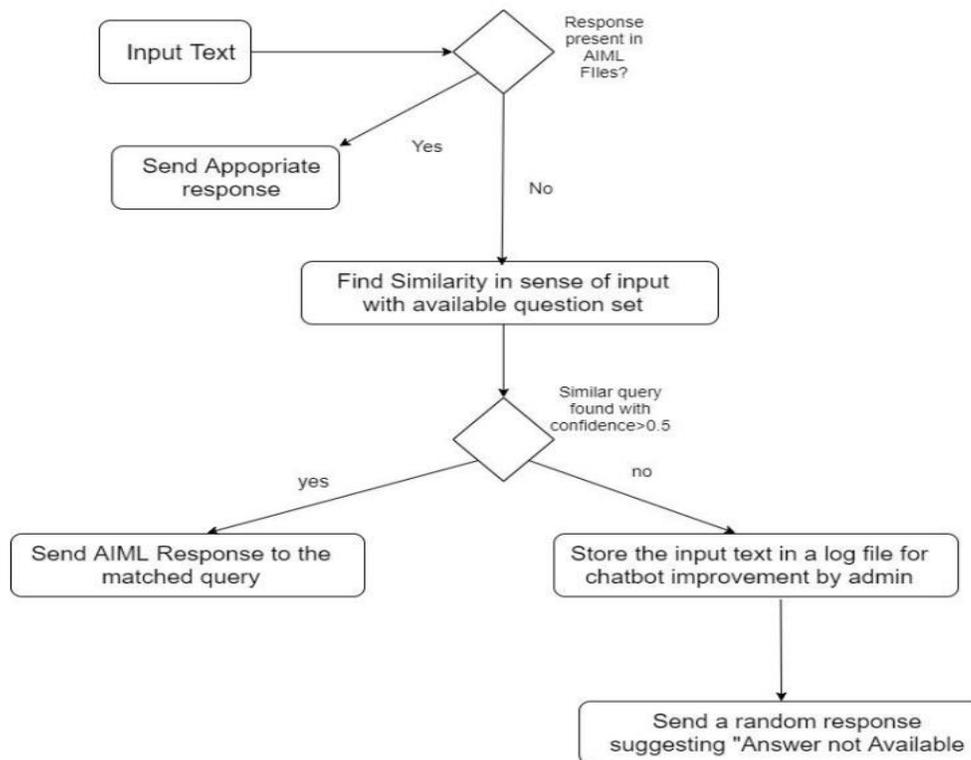


Fig. 1: SYSTEM FLOWCHART

IV. ALGORITHMS

A. NATURAL LANGUAGE PROCESSING (NLP)

Natural language processing (NLP) is a subject of computer science—specifically, a branch of artificial intelligence (AI)—concerning the ability of computers to understand text and spoken words in the same manner that humans can. NLP is implemented using a library in Python named NLTK. The process of teaching computers to interpret natural language is known as natural language processing. Computers can interpret structured data such as spreadsheets and database tables, but unstructured data such as human languages, words, and voices is difficult for computers to comprehend, necessitating the use of Natural Language Processing.

There are some steps followed in NLP:

Tokenization: It is the process to break a complex sentence into words. Also, the importance of each word is understood with respect to the sentence.
Stemming: It is the process in which words are normalized into its base form or root form.
Lemmatization: It is the process in which grouping of different inflected forms of a word is done. It also roots several words into one common root but the output of Lemmatization is a proper word.
Stop Words: These are some of the words which are helpful to make a sentence meaningful but do not help in NLP.

Parts of Speech: It is an inbuilt library containing the various parts of speech. For e.g.: 12 CD- Cardinal Number NN- Noun Singular, NNS- Noun Plural; etc.

Named Entity Recognition: It aids in the identification of a certain entity's name. For example, a movie, its monetary value, its organisation, its location, or its quantity.

Chunking: It is a process of picking up individual pieces of information and grouping them.

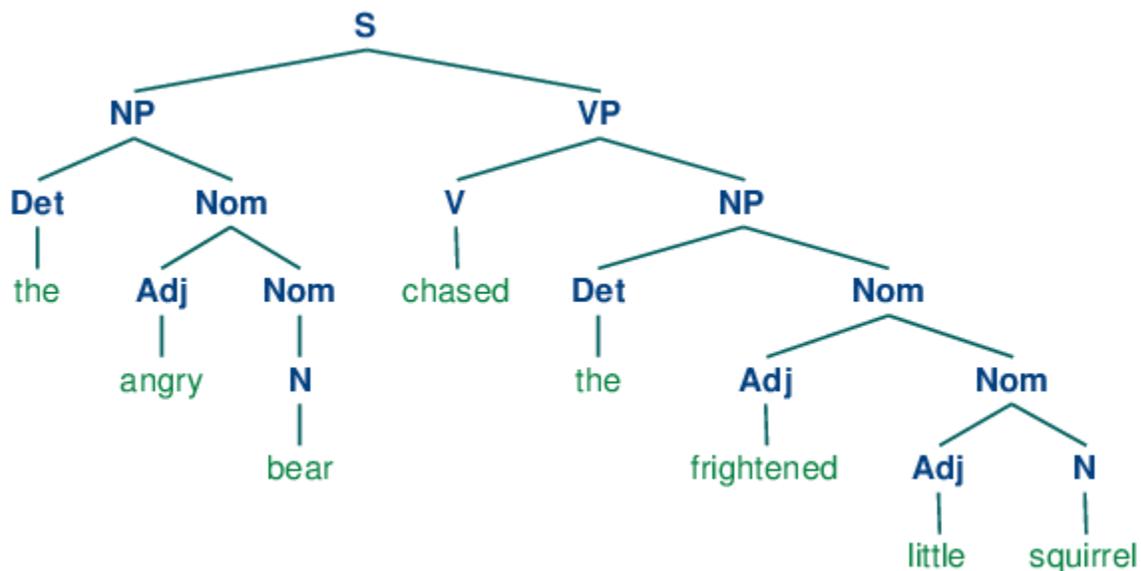


Fig. 2 : Syntax Tree – Natural Language Processing

B. PATTERN MATCHING ALGORITHM

In chatbots, pattern matching was one of the most commonly employed algorithms. The Pattern Matching Algorithm is a database that comprises questions and answers. Patterns are used to name questions, while templates are used to name replies. Artificial Intelligence Mark-up Language (AIML) tags make up the response to this specific query. Patterns and templates are kept in a tree format. Questions are on the branches, and responses are at the nodes, thus anytime a user asks a question, the query is first searched for an answer word by word, and then the specific answer is fetched from the node. The

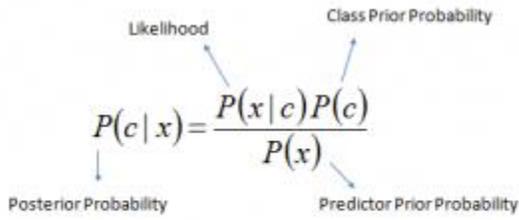
ALICE chatbots employ this type of framework. The advantage of this method is that it allows users to quickly obtain answers to questions that have already been stored. As a result of its simplicity, it is extensively used. This algorithm only retains certain types of questions, so if the user asks a question that isn't on the list, it won't be able to respond. As a result, it is incapable of self-learning.

C. NAIVE BAYES ALGORITHM

Chatbots can be built using the Naive Bayes approach in which we just have to give each training question a corresponding label number of the answer and the classifier will predict the label of the user input. This label number is an index of a corresponding response in our answers' dataset. The Bayes' Theorem is used to generate the Naive Bayes classifiers, which are a collection of classification methods. It is a family of algorithms that share a similar idea, namely that each pair of features being classified is independent of the others.

It's a classification method that uses Bayes' Theorem and assumes predictor independence. A Naive Bayes classifier, to put it simply, assumes that the presence of one feature in a class is unrelated to the presence of any other feature. For example, If a fruit is red, round, and roughly 3 inches in diameter, it is believed to be an apple. Even if these characteristics are reliant on one another or on the presence of other characteristics, each of them contributes to the likelihood that this fruit is an apple, which is why it is called 'Naive'. The Naive Bayes model is simple to construct and is especially good for huge data sets. Naive Bayes is renowned to outperform even the most advanced classification systems due to its simplicity. It is simple and quick to forecast the test data set's class. It's also good at multi-class prediction. When the assumption of independence is met, a Naive Bayes classifier outperforms alternative models such as logistic regression and requires less training data.

The Bayes theorem allows you to calculate posterior probability $P(c|x)$ from $P(c)$, $P(x)$, and $P(x|c)$ using $P(c)$, $P(x)$, and $P(x|c)$. Look at the equation below:


$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$

Above,

$P(c|x)$ is the posterior probability of class (c , target) given a predictor (x , attributes).

$P(c)$ is the prior probability of the class.

$P(x/c)$ is the probability of the predictor given class.

$P(x)$ is the prior probability of the predictor.

V. CONCLUSION

The system's purpose is to keep students informed about their college activities. With the use of Artificial Intelligence and Knowledgeable databases, Artificial Intelligence is the fastest developing technology on the planet. Pattern matching and virtualization are two areas where we can make changes. This system is creating a conversation bot based on the Android operating system, which combines Artificial Intelligence, Knowledgeable Databases, and Virtual Assistance. We can develop such a chat bot which will make a conversion between human and machine and will satisfy the question raised by the user. The project's major goal is to reduce the amount of work that the college's office staff has to do.

REFERENCES

1. <https://sci-hub.se/https://ieeexplore.ieee.org/document/8692927>
2. <https://sci-hub.se/https://ieeexplore.ieee.org/document/9182168>
3. <https://www.ijert.org/research/college-enquiry-chat-bot-system-IJERTV9IS090396.pdf>
4. https://www.ijntr.org/download_data/IJNTR03010029.pdf
5. M. J. Pereira, L. Coheur, P. Fialho, and R. Ribeiro, "Chatbots' greetings to human-computer communication," arXiv preprint arXiv:1609.06479, 2016
6. O. Deryugina, "Chatterbots," *Scientific and Technical Information Processing*, vol. 37, no. 2, pp. 143–147, 2010.
7. J. S. Malik, P. Goyal, and A. K. Sharma, "A comprehensive approach to chatbot"
8. B. A. Shawar and E. Atwell, "Chatbots: are they really useful?" in LDV
9. A. Bordes, S. Chopra, and J. Weston, "Question answering with subgraph embeddings," arXiv preprint arXiv:1406.3676, 2014
10. B. Setiaji and F. W. Wibowo, "Chatbot using a knowledge in database: Human to-machine conversation modeling," in *Intelligent Systems, Modelling and Simulation (ISMS)*, 2016 7th International Conference on. IEEE, 2016, pp. 72–77.
11. H. Wang, Z. Lu, H. Li, and E. Chen, "A dataset for research on short-text conversations." in *EMNLP*, 2013, pp. 935–945.
12. I. V. Serban, A. Sordoni, Y. Bengio, A. C. Courville, and J. Pineau, "Building end-to-end dialogue systems using generative hierarchical neural network models." in *AAAI*, 2016, pp. 3776–3784.
13. J. Hill, W. R. Ford, and I. G. Farreras, "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations," *Computers in Human Behavior*, vol. 49, pp. 245–250, 2015.
14. N. Asghar, P. Poupart, J. Xin, and H. Li, "Online sequence-to-sequence reinforcement learning for open-domain conversational agents," arXiv preprint arXiv:1612.03929, 2016.
15. I. Ahmed and S. Singh, "Aiml based voice enabled artificial intelligent chatterbot," *International Journal of u-and e-Service, Science and Technology*, vol. 8, no. 2, pp. 375–384, 2015.