

## **“CHATBOT FOR SSGMCE WEBSITE”**

<sup>1</sup>JIGAR WAKHARIYA

Department of Information Technology, Shri Sant Gajanan Maharaj College of Engineering, Shegaon Maharashtra, India  
jigarwakhariya@gmail.com

<sup>2</sup>PRAKHAR GANGRADE

Department of Information Technology, Shri Sant Gajanan Maharaj College of Engineering, Shegaon Maharashtra, India  
prakhargangrade28@gmail.com

<sup>3</sup>PUSHKAR JAODAND

Department of Information Technology, Shri Sant Gajanan Maharaj College of Engineering, Shegaon Maharashtra, India  
ppjaodand@gmail.com

<sup>4</sup>ASHISH BHAGWAT

Department of Information Technology, Shri Sant Gajanan Maharaj College of Engineering, Shegaon Maharashtra, India  
ashishbhagwat2015@gmail.com

<sup>5</sup>AMITKUMAR MANEKAR

Department of Information Technology, Shri Sant Gajanan Maharaj College of Engineering, Shegaon Maharashtra, India  
asmanekar24@gmail.com

**ABSTRACT:** Chatbots are used widely for interactive communication between machine and humans from a very long time. Although the technology and application of Chatbot has changed with time. Chatbot is generally a virtual person or assistant who can effectively talk to any human being using interactive textual skills provided to it. The motivation behind writing down this paper is that it will be helpful for both Professors, visitors and students to ask any sort of questions related with the Shri Sant Gajanan Maharaj College of Engineering and to comprehend rationale behind this. Our special importance is actually based on accuracy of determining chatbot system. However, the technology enables people for dealing with machine in their language by means of a machine interface is picking up prominence in an assortment of questions mainly for benefit of user. The complete model of an algorithm trained with wide dataset, with interactive UI's which are responsive to various classified actions make the experience of users amazing. The chatbot system will be built using tools from Keras library of Python. The response principle matches the input sentences from a user. The User can enter their queries and doubts in the textbox of chatbot to get an appropriate output. Thus user can get their queries solved virtually without visiting the enquiry counter. The System analyses the query and then provide a response to the user.

**Keywords:** Chatbot, NLP (Natural Language Processing), NLTK( Natural Language toolkit), Pattern matching approach

### **1. INTRODUCTION**

Chatbots are conversational interfaces as they are additionally known, provide a new way for people to collaborate with machine frameworks. Generally for getting an query replied by a software program included utilization of an internet searcher, or filling out a form. A chatbot system enables a person to ask queries in the similar manner that they would address a human being. The most surely understood chatbots as of now are voice chatbot system: Alexa and Siri. In any of the case, chatbots are right now being embraced at a high rate on machine visit stages. Later In machine learning have significant enhanced the precision and maintain adequacy of Natural Language Processing by making the chatbots a practical alternative for numerous associations. This improvement in Natural Language Processing is initiating a lot of extra research which should proceeded with a difference in the viability of chatbot system in the years to come. The chatbot replies using an effective GUI (Graphical User Interface) providing a real experience as such as someone is really listening and responding humanly.

In this project, we created a chatbot which will provide all the information regarding Shri Sant Gajanan Maharaj College of Engineering through the web interface using Deep Learning and NLP techniques. Thus, our bot will act as a person at enquiry office whom they can ask any questions regarding admission and college environment at any time in more interactive way. Thus,

reducing workload of the admission cell personnel. This system can be integrated on college website and thus students who desire to get admission in the college can clear their doubts from their home. We have provided a feedback mechanism through which user can rate bot's responses and thus bot can respond more accurately next time. The KNN algorithm is used to solve the classification model problems and basically creates an imaginary boundary to classify the data. When every new data points come in, the suitable algorithm will definitely try to predict that to the nearer of the boundary line. This bot can maintain context and can give responses based on the context by operating according to the algorithm. Thus, this paper focuses on the sample bot which can be used as a reference for creating chatbots for many other purposes. We used the Tokenization process of replacing sensitive data with unique identification symbols that retain all the essential information about the data with more security which converts the normal strings of text into a list of tokens i.e. words that we actually want. Sentence tokenizer will be used here to find the list of sentences and we used tokenizer called as Word tokenizer which will be used to find the list of words in strings. The paper by explaining the working of this sample chatbot explains the proposed methodology of developing similar purpose chatbot.

## **2. LITERATURE SURVEY**

The authors in [1] describe an idea to determine several significant facts in the text representing the life of a historical figure to build a complete conversational chatbot. This Chatbot should be able to learn (supervised learning) from previous experiences in order to respond more realistically. The authors provide a generic form of sentences to solve the issues and problem of learning to enable the Chatbot to acquire as much crucial information as possible relating to the particular person and life of the person being simulated. These chatbots get such information from websites such as: The unstructured data obtained from Wikipedia and DBpedia for structured data.

NLP techniques are used to convert plain text to structured text and then restructure them into a generic form of a sentence [1]. The input fed to the chatbot is a collection of certain factual words which are transformed in a way such that they cannot match the generic form. The authors used an open source Chatbot called ChatScript to design the conversation. ChatScript uses only a simple word (lexical) matching technique to search for the answer. In [4] the authors explain the supervised discriminative models which learn to choose or rank and classify answers by using examples of question, answer pairs. The representation of Question Answer (Q/A) pair is provided via kernel combinations applied to its members. To reduce the load of large amounts of manual annotation, the authors represent Q/A pairs by means of generalization methods, employing the application of structural kernels to syntactic/semantic structures [4]. This proposal uses more than one matching metric in order to find a better match to the query. The same kind of data as in [1] is used in the proposed system. We combine semantic, lexical, and syntactic features extracted from a text as in order to obtain an optimized match to a query by quantifying the output results [5]. The contributions are presented in this paper: the first contribution is populating an empty knowledgebase of a new born Chatbot by sentences automatically from the web according to user choice for a figure or an object to simulate by the Chatbot. The second contribution is using a new combination of multiple feature extraction methods to find a best response to a Chatbot query. [1] Question Answering (QA) systems can be identified as information accessing systems which try to answer to natural language queries by providing answers instead of providing the simple list of document links. QA system selects the most appropriate answers by using linguistic features available in natural language techniques. They differ mainly from various knowledge sources, the broadness of Dialog Systems (NLDS) is a suitable and easy way to access information. Two approaches are that simplifies chatbots realization. The first approach is Ontology, which is represented in a twofold manner: to construct answers very actively as a result of a deduction process about the domain, and to automatically approach, off-line, the chatbots KB with the words that can be derived from the ontology, describing properties and relations between concepts involved in the conversation. Second is to preprocess the sentences given by the user so that it can be reduced to a simpler form that can be matched to existing queries of the chat-bots. The aim is to provide useful information regarding their interests and to obtain what they desire. The choice was to implement a QA system using a typical pattern-matching chatbots technology.

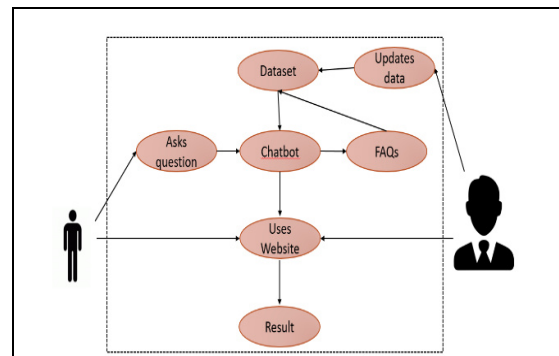
[2] This paper describes an approach to the idea of determining the most important facts in the texts describing the life of a historical figure for building an interactive

conversational agent.

Chat-bots are mainly used to simplify the conversation between the human and machine [3]. Admin feeds certain knowledge in the form dataset to the machine so that machine can determine the sentences and taking a decision itself as a response to answer that question. The chat used in this paper is actually in the form of Indonesian conversational pattern and MySQL is used as database in this project. It can miss defining a sentence and how to respond it while connecting chat application to the database. So the knowledge representation and application of SQL in the pattern-matching operation are needed. A data that has been modeled based on the pattern of the conversation would be then tested by the help of a series of conditions. The chatbot conversations are cross checked to obtain accuracy. It is done so that it can enhance the knowledge of the database by adding certain knowledge as it has not been modeled before. If in case the input data in the database did not match with them then it will be remodeled.

## **3. METHODOLOGY**

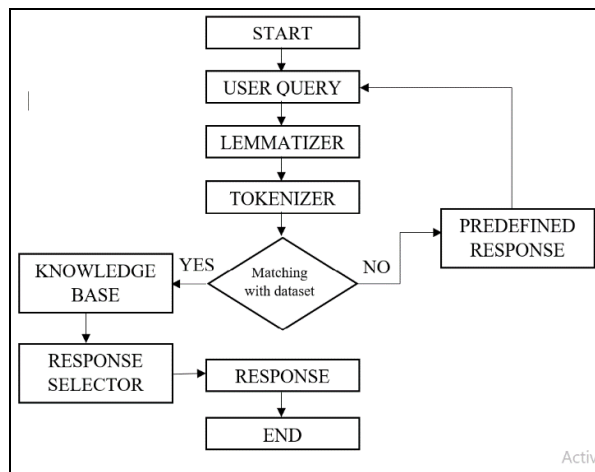
This System is a web application which gives reply to the question of the user. The student or any user can question about the college related exercises and activities through interacting with the chatbot. The basic Use Case of the chatbot is shown in Fig.1. We used Keras Neural Network which uses the entire dataset in its training phase. When a proper prediction is necessary for an unseen data instance, it directly searches through the whole training dataset for the most similar instances and the data having most similar instance is then returned as the final prediction.



**Figure 1: Use Case Diagram**

We used NLTK data package which includes pre-trained tokenizer for English language which removes noise (which is not standard number or letter) and removes the common words which would appear to be of small values in helping select documents, for matching a needs of users are get excluded from the vocabulary entirely.

The chatbot system will simply take the queries of the users which can be anyone, and will put responses in accordance with the queries asked by the user. We have used a Lemmatizer function by NLTK that reduces the extra words e.g. “developer”, “developed”, “develops” reduce to the stem “develop”. Lemmatization is very similar to stemming but it brings context to the words. So it links the given words with another word with same meaning.



**Figure 2:** System Architecture

The system architecture of the chatbot shows complete process flow of the system which includes the matching of dataset with the input query. This query is applied with nltk tools Lemmatizer and Tokenizer and then matched with dataset to obtain probable responses which are then evaluated under the algorithm to get the accurate result. The algorithm numerically assigns probabilities to the different words and the highest probability index is then obtained as the response as shown in Fig.2.

The use of the chatbot system is to match the queries of the user with knowledge base and see for the proper appropriate response. The system can also reply to the general queries of the student. The algorithm of the complete system can be understood by the following flow chart. The proposed system consists of three main parts which are as follows:

- 1) User Interface
- 2) Neural network model and NLP unit
- 3) Feedback System

### 3.1 User Interface

The user interface can be in the form of anything from desktop to phone and web. The interface in the example is a web interface. The interface will also work as the instruction guide for the users. The user will get all the information needed to understand the system in here. The main purpose of the interface will be to take the queries and convey them to the backend which consist of tensor module and feedback system. The interface works as a medium between the backend and user.

### 3.2 Neural Model and NLP Unit

This module is the main core of the whole system. This is the part which generates the actual response for the user query. First the model is defined using TensorFlow which is trained using intent file created. Intent file is in form of json file which is as follows:

```
{ "tag": "admission", "patterns": ["tell me regarding admission procedure", "what is admission process", "enquiry about admission"], "responses": ["Admissions are closed", "You are little late, admissions have just closed", "Wait for a year and you will get admission next year"] },
```

```
{ "tag": "application date", "patterns": ["when do i apply?", "What is the due date for application?", "what is last date of the application?", "what is starting date of the application?" ], "responses": ["Application process will remain open at 04th of July to 14th of august", "from 4th July to 14th august", "you can apply from 4th july to 14th august"] },
```

```
{ "tag": "Location", "patterns": ["where is the location of the college?", "College is located at?", "How to reach college?"], "responses": ["shri sant gajanan maharaj college of Engineering is located near khamgaon road sheagaon ."] },
```

It consist of three parts i.e. Tags, Patterns, Responses. The tag is nothing but context of that query. It defines what query is about. Patterns and responses as name suggests is used to train the model with sentences and get corresponding responses. Patterns are loaded and passed through query modulation.

We used a Python flask framework for making web app ,make directories for templates and train the bot also used Bag of words algorithm which is concerned with whether the known words occur in the document, not where they available in the document and transform the text into a meaningful vector (or array) of numbers. Now when user gives query this query will be passed through same query modulation process and converted into bag of words. This will be given as input to model and model will return all the tags that have probabilities greater than error threshold defined. Then response model will select the response which will be more accurate and having greater ranking given by feedback System.

### 3.3 Feedback system

The database stores the intent file which is crucial part of the training and response. User interface will be provided with functionality through which user can rate the response and this will be stored in database along with responses. Next time when user gives query, response model will give response based on the feedback i.e. this response will be checked for ranking and highest-ranking response will be displayed. This ranking will be integer based with 5 having highest ranking. Initially all the responses will be set to default 0 and as and when user interacts with the bot it will be updated.

## 4. DISCUSSION OF CURRENT SCENARIO

- If Students need to visit manually to the college to get their queries answered by the college help desk.
- This process need lot of time as well as money as the customer needed to visit college if its miles away from home.
- Also, this process may lead to gap in communication between the student and college.
- No LIVE update platform available yet for all student and parent queries
- No proper alternatives available for student and parents beyond visiting administration of college.

## 5. CONCLUSION

The Chatbot serves as a single solution to the multiple problems, it is far better to build a chatbot rather than a Search Box. The interactive behaviour of chatbot enhances the user experience. Thus chatbot could be considered as the best solution for the problem of automating the enquiry system and query processing.

The chatbot technology is evolving day-by-day making chatbots smarter and accurate. In near future, it will be possible to get chatbot trained without supplying any type of data but by the active neural networks and applying human observation AI techniques which is a very wide area of research and scope. But there is a less probability of the change in the principles.

## **6. REFERENCES**

- [1] E. Haller, and T. Rebedea, "Designing a Chat-bot that Simulates an Historical Figure." pp. 582-589, 2013.
- [2] Emanuela Haller, Traian Rebedea Faculty of Automatic Control and Computers university Politehnica of Bucharest, 978-0-7695-4980-4/13 \$26.00 © 2013 IEEE. "Designing a Chat-bot that Simulates a Historical Figure".
- [3] Bayu Setiaji, Ferry Wahyu Wibowo , Department of Informatics Engineering STMIK AMIKOM Yogyakarta, Yogyakarta, Indonesia, 2166-0670/16 \$31.00 © 2016 IEEE "Chatbot Using A Knowledge in Database-Humanto-Machine Conversation Modeling".
- [4] A. Moschitti, and S. Quarteroni, "Linguistic kernels for answer reranking in question answering systems," Information Processing & Management, vol. 47, no. 6, pp. 825-842, 2011.
- [5] N. Kambhatla, "Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations." p. 22, 2004.