



CHATBOT USING NATURAL LANGUAGE PROCESSING

¹Pushkar Shinde, ²Priyanka Boraste, ³Sunil Datir

MVPS's KBT COE , Nasik

¹shinde.pushkar@kbtcoe.org, ²oraste.priyanka@kbtcoe.org, ³datir.sunil@kbtcoe.org

Abstract— Chatbot is an application where the computer talks with the user like an human. The software can understand the discussion it is having with the user and respond accordingly. Chatbot uses the concepts of Natural Language Processing (NLP) and Artificial Intelligence to achieve this task. Chatbots can be useful in various applications. It can be used as a support source for answering questions in various domains, it can be used as a guide during learning languages and in a User Interface Agent for applications. The knowledge base is an important part of an Chatbot application. The knowledge base should store all the information of the domain in which the Chatbot is operating. It should be able to provide information to the application to answer questions asked by user. Chatbot used as a guide in language learning can correct the grammatical mistakes done by user in communication with chatbot. The Chatbot will have knowledge stored of grammatically correct sentences. On communication of the user with Chatbot, it will check the correctness of the statement and inform user.

Index Terms— Artificial Intelligence, Chatbot, Natural Language Processing.

I. INTRODUCTION

Artificial Intelligence is an field in Computer Engineering in which computer behaves like a human. Natural Language Processing is an field in Artificial Intelligence where computer understands the human languages like English, Marathi, Hindi etc. Chatbot is an application of Natural Language Processing where user can chat with the computer like a user would have otherwise done with a human. Using NLP the

users can communicate with computer using human languages. In an Chatbot the user would enter textual information in human language like English which the computer will understand and will respond back in English,.

Language learning Chatbot can be used by user for language learning. The user- will communicate with computer in English and the computer will tell user's spelling mistakes and grammatical mistakes. This kind of chatbot can be useful for users who want to practice English speaking with an virtual assistant. A teacher can't be available 24x7 to an student when student is doing his speaking practice, so in that scenario this Chatbot can be useful. TextGear[2] API is a freely available linguistic resource. Word segmentation, semantic mapping, and syntactic parsing can be done by TextGear API. Grammatical and spelling mistakes in the user's sentence can be found by TextGear and users are provided with appropriate grammar structure.

Levenshtein distance [2] can be used for finding similarity. Relatedness between sentence pairs can be found by Lavenshtein distance which is an NLP task. It finds if the two sentences are paraphrased or not. There can be a knowledge base as an question answer database which saves a statement and corresponding answer. PhpMyAdmin can be used to store the databases on the localhost. MySql databases can be handled using PhpMyAdmin, allowing developers to perform complex operations using its UI. Porter stemming algorithm [3] (or 'Porter stemmer') is a process which removes suffixes from words in English. This algorithm can also be used in an Chatbot applications for formatting sentences. Procedure for Paper Submission

II. RELATED WORK

A. Chinese question Answering and Answer Validation Web Application

A web-based Chinese automatic QA system uses an answer extraction method which is based on the calculation of sentence similarity between question and answer [4]. The similarity between question and answer for a web based Chinese automatic QA system is an example of a chatbot used [4]. It uses calculation of sentence similarity. It understands the question type and answer type by analysing the question. The best answer is chosen by parameter of similarity and if the answer is found the procedure ends.

B. Natural Language survey for grammar checkers

If the sentence given to the Chatbot is grammatically incorrect the input is corrected in the output or suggestions are given related to the input [5]. If there is a grammatical mistake this approach suggests the user about the grammatical mistake or replaces the incorrect word with correct word.

III. METHODOLOGY

A. Chatbot for language learning

The English language is mostly used in the world. People want to master it because it is the language of international communication. Most of the traditional ways of learning the language include books, in-person coaching classes even online coaching. The drawback here is the necessity of availability of a tutor, the need to follow certain timings, lack of guarantee of this tutor is as knowledgeable as he claims, and lack of fun in the process. To learn new things effectively we must be in the mental comfort to process and analyze new information. Also to learn a new language we must practice in that language as much as possible. Forming sentences without any context can make the learner lose interest.

Nowadays, Chatbots are used in many applications, one of which is language learning. It is found that users' response to a virtual friend is much more positive when they are trying to learn new things. They converse freely without the fear of social stigma. Also learning using a chatbot in one's time of ease makes learning effective. To address all these drawbacks of traditional learning and explore the advantages

of a chatbot we propose a new system that can act as a chatting partner as well as a tutor to the user [2]. The system will be an android application which will provide two modes to the user. First, normal chat, wherein the app will provide output to the user's query without any grammar check. Second, chat with a linguistic and grammar check. Here, the system will process the user's input and check semantic, linguistic, and vocabulary errors in the statements. Thus, a user can use the tutor mode of our app when he/she wants to learn and a chatting partner mode otherwise. The availability of both of these modes makes sure that the app provides both entertainment and educational value to our users. Most of the chatbots developed work on the procedure of recognizing patterns in the input sentence and responding from a template. Research on the past chatbots such as A.L.I.C.E has shown that users are put out by response from templates as this most of the time are unrelated responses. Another method is a keyword-based matching system. The drawback of this is that it does not take into account the semantics and relationship between the words in the original query. We have used another approach; we have tried to find a similarity between two sentences rather than the pattern. In this approach, we use Levenshtein distance for finding similarity. It finds relatedness between sentence pairs which is an NLP task. It finds if the two sentences are paraphrased or not [2]. They have used a question-answer database which saves a statement and corresponding answer. PhpMyAdmin is used to store the databases on the localhost. It is a freely available tool to handle MySQL databases allowing developers to perform complex operations using its UI. The app is designed to emulate the look and feel of modern-day popular messaging apps such as messenger and WhatsApp. This ensures minimum user preparation for working with the system. We have also provided the option of speech output. This would help in improving the dialect of the user. For grammar and vocabulary checking we need wide vocabulary and linguistic rules building which is complex and can decrease accuracy. TextGear API is a freely available linguistic resource. TextGear performs word segmentation, semantic mapping, and syntactic parsing. It would find grammatical and spelling mistakes in the user's sentence and

provide appropriate grammar structure for our users.

In one of the proposed systems of Chatbot there is an 2 tier architecture. In this system there is an retrieval based model. The system is using PHP with XAMPP server and MySQL JSON format is used for network requests between the app and the database. For connectivity to the app Android Volley is used. The Volley network works asynchronously. The app uses in-build Google speech to text and text to speech API for communication in speech format [2]. The chatbot can help the user to do practice of English language learned. The user can communicate with the chatbot in English and the chatbot can not only communicate with the user but also check the grammatical mistakes made by him during the chatting. When chatbot points out the mistake made by him, it will help the user to understand his mistake and in turn correct his mistake. The chatbot uses the knowledge for chatting with the user and for identifying the grammatical mistake made by the user during chatting. This will be very helpful resource for both Teacher and student, as once the teacher teaches a feature of English language to the student, the chatbot can do the practice of the English learnt by the student. The knowledgebase is created in MySQL and supports the chatbot.

For the detection of grammatical mistake done the chatbot will find the type of the mistake made in chat. The type will be searched in the knowledgebase and the mistake will be found. Once the type of the grammatical mistake is found the chatbot will inform the mistake to the user. [Figure 1] represents a complete system architecture.

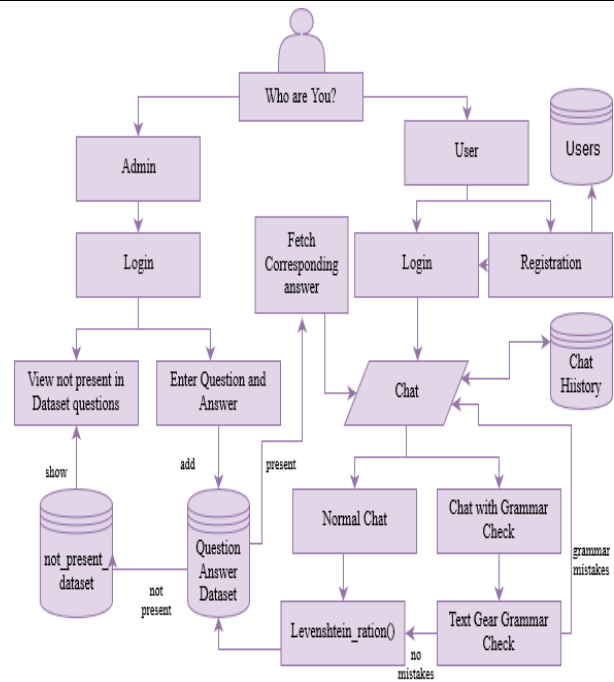


Fig. 1. System Architecture [2]

Algorithms used for this project are as follows:

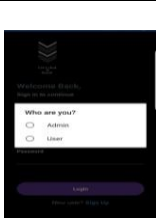
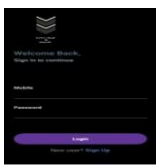
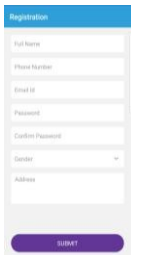



User Access


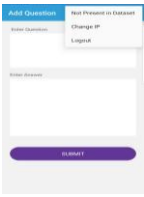
- Step 1. Select user role.
- Step 2. Register for a new user account.
- Step 3. Login to the account using username and password.
- Step 4. Ask question to the Chatting Partner Application of your interest via: 30 speech input or typing it directly using keypad.
- Step 5. Select the chatting mode:
 - Normal chat, the system will respond to the query with respond to the predefined answers.
 - Check for the grammar errors and get correct answer.
- Step 6. Report, to add your question to the dataset.
- Step 7. Logout from the application.

Admin Access Algorithm

- Step 1. Select admin role.
- Step 2. Login with admin pre-defined login credentials.
- Step 3. Update the dataset with new questions and answers or add use reported questions and their answers to the dataset.
- Step 4. Logout from the application.

Table 1 Implementation and Results [2]

1	There are two defined roles with which the user can log in, Admin and User	if (\$user_role=="1") // echo "success_admin"; else if (\$user_role=="2") //echo "success_user";	 Figure 2
2	User/Admin login	select * from user where credential slike '<username>' and password = '<encrypted password>'	 Figure 3
3	Registration Activity is ONLY available for users	Required field: \$name= \$_POST['name'] \$phone = \$_POST['phone'] \$password=\$_POST['password'] \$email = \$_POST['email'] \$address = \$_POST['address'] \$gender = \$_POST['gender']	 Figure 4
4	Chat Activity	The chat data is displayed on the user screen using RecyclerView	 Figure 5
5	Chat features a)Chat without grammar correction b)Chat with grammar correction		 Figure 6
6	Grammar detection and correction	curl_setopt(\$ch, CURLOPT_URL,"https://api.textgear.com/check.php"); curl_setopt(\$ch, CURLOPT_POST,1); curl_setopt(\$ch, CURLOPT_POST_FIELDS, http_build_query(array('text' =>	 Figure 7

		\$text,'key'=>\$key');));	
7	Fetch answer to users question	select answer from question_answer where levenshtein_ratio(' \$sender_question', question) >= '90'OR select answer from question_answer where levenshtein_ratio(' \$sender_question', question) >= '80	 Figure 8
8	Admin activity Add data to the database	Get data into „nameValuePair“ HttpResponse response = httpClient.execute(httpPost);	 Figure 9

In case the user makes multiple mistakes made by the user the chatbot will detect the multiple mistakes from the knowledgebase and inform the multiple mistakes to the user. Lavensthein distance is difference in words of an sentence with words of other sentence stored in the knowledgebase. The sentence with least Lavensthein distance will be selected from knowledgebase.

The Lavensthein distance between words of two sentences a , b (of length |a| and |b|) is given by lev(a, b) where

$$\text{lev}(a, b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \text{lev}(\text{tail}(a), \text{tail}(b)) & \text{if } a[0] = b[0] \\ 1 + \min \begin{cases} \text{lev}(\text{tail}(a), b) \\ \text{lev}(a, \text{tail}(b)) \\ \text{lev}(\text{tail}(a), \text{tail}(b)) \end{cases} & \text{otherwise.} \end{cases}$$

Fig. 3. Lavenshtein distance [2]

B. Chatbot for College Enquiry

Another Chatbot application suggested is the college inquiry chat-bots will be built using Natural Language Processing algorithms. It analyses user's queries and understand user's message. Matching the input sentence from a user is the response principle in this app. Questions can be asked to the Chatbot for college related activities without coming physically to the college. The User can ask the question any college-related activities through the chat-bot without physically available to the college for inquiry. The answers to the questions are obtained by the system by analysing the questions. The system answers the questions by the users with the help of Natural Language Processing. The system replies using an effective Graphical User Interface as if a real person is talking to the user [6] User registers himself/herself on Chat-Bot application. Then submits his/her complaints and queries regarding college.

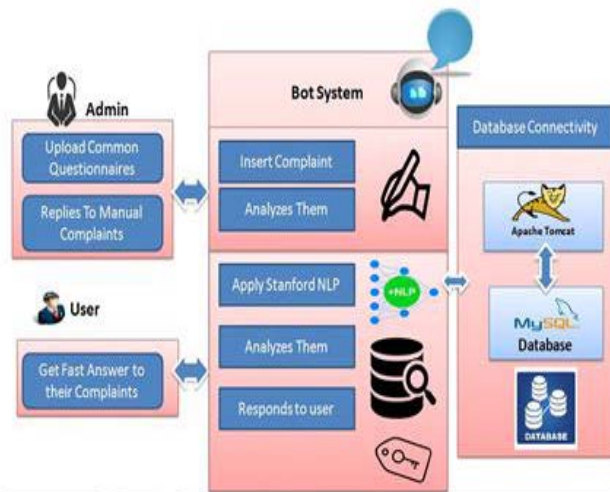


Fig. 2. System Architecture [6]

The knowledgebase for the college chatbot is an important part of it. The knowledgebase stores the information required by the chatbot to talk with the user. The college chatbot knowledgebase stores the complete information of the college to chat with the user. The user can

ask required information to the user and the chatbot will answer the question using the information stored in the knowledgebase. The chatbot can act as an answering system 24 x 7 for different users of the college website. The college chatbot would be deployed on the college website. The chatbot will search the answer in the knowledgebase of the question asked by user. On finding the answer to the question, the chatbot will give the answer as a reply. The more the information in knowledgebase of the chatbot the better will be the response given by chatbot.

IV. CONCLUSION

Thus we have seen in this paper the various applications of Chatbot. Chatbots can help in answering questions of users by using the knowledge base. The Chatbot responds to the questions of users using Natural Language Processing. The Chatbot behaves like a human when discussing with the user. This helps the organizations to deploy an 24x7 support for College inquiry and also in application like Language learning.

REFERENCES

1. H. N. Io, C. B. Lee, "Chatbots and Conversational Agents: A Bibliometric Analysis", 2017, IEEE, China.
2. Kunti Sandanshi, Ashwini Bande, Samruddhi Jumale, Pushkar Shinde "Chatting Partner for English Language Learners", IJIRT, vol. 6(2020)
3. K. Bala, Mukesh Kumar, Sayali Hulawale, Sayali Pandita "Chatbot for College Management System Using A.I." , IRJET, vol. 4(2017)
4. Dongfeng CAI, "A Web-Based Chinese Question Answering Validation", IEEE International Conference on Natural Language Processing and Knowledge Engineering, China(2005).
5. Nivedita Bhirud, "A Survey Of Grammar Checkers For Natural Languages", 3rd International Conference on Signal, Image Processing and Embedded Systems, India(20170).
6. Jiyou Jia, "CSIEC(Computer Simulator in Educational Communication): A Virtual Context-Adaptive Chatting Partner for Foreign Language Learners", IEEE International Conference on Advanced Learning Technologies., Germany(2004).