Approach for Virtual Translator using Natural Language Processing

[1]Hamsaveni M
Prof., Department of CSE
VVCE, Mysuru

[2]Eshwar Thilak R
Department of CSE
VVCE, Mysuru

[3]Anusha U
Department of CSE
VVCE, Mysuru

[4]Gowthami C P
Department of CSE
VVCE, Mysuru

[5]Avantika Holla S
Department of CSE
VVCE, Mysuru

Abstract:- Dissemination of information demands the presence of at least two people. Sometimes, due to the absence of one person information is lost which would lead to closed loops. This paper proposes an application of machine learning for the educational sector, rendering a particular solution to the problem of loss of information. Various techniques such as speech recognition, Natural Language Processing (NLP), tokenization and stemming are used in this system.

Keywords:- Tokenization, Stemming, Natural Language Processing (NLP), Natural Language Toolkit (NLTK)

I. INTRODUCTION

For an institution to run in an organized manner it is necessary to facilitate a helpdesk that would minimize the confusion a person has with his queries. This helpdesk is usually run by a dedicated authority who is present throughout the day to provide information when required. Absence of a receptionist might result in chaos and hence introducing a virtual receptionist would minimize this issue. This possesses an ability to provide response to all the queries and enhances the automation of day-to-day tasks. This leads to increased productivity, better quality of service and improved user satisfaction

II. EXISTING SYSTEM

Most of the universities, educational institutions, hospitals, banks, all have a receptionist dedicated to answer the queries and other details that a person needs to be clarified. This necessitates a person to be present at the front desk to give the information needed. Take for example, in an institution there is always a front desk made available and the receptionist has to be present throughout the day so that the information is reverted accordingly. During counselling or admissions, the influx of queries that has to be answered is huge and absence of a person in the front desk will cause numerous problems. There is a possibility of information being lost due to the fact that a receptionist is not present every single minute of the day and there is no other person to replace or cover for his/ her absence, or the details delivered might not be satisfactory. Since customer satisfaction determines or forms the base for successful running of an institution, there are chances of losing people's interest as a result of this which would ultimately affect the institution. Hence we have proposed a system that will overcome these issues caused.

III. PROPOSED SYSTEM

In the existing system, communication becomes a difficult task due to the absence of a person that causes delay of information being received at a particular time or even a loss of information. To overcome this problem, we propose a method which makes use of natural language processing for generating user satisfactory information to convert speech to text and give an appropriate reply in both speech and text. This approach seems to be interesting because the user or a person might have numerous queries that needs to be cleared and hence can rely on the proposed system.

The proposed system makes use of natural language processing toolkit (nltk) wherein a speech utterance by a person is taken and is categorised into words which are being put into a process of stemming, tokenizing, pattern matching with data that is already fed into the system to provide required response. This system makes uses of database which holds the information of the users, their queries, and also date and time during which they were asked. In our system, the principal module is connected to the database that holds the user information. The proposed system is represented in two steps:

➢ *Step 1: The conversion of user speech to text*

Here the input speech is captured by the microphone which is then recorded and eventually converted into text. The text that is recorded is converted into tokens. The method word_tokenize() is used to split the sentence into its component words. From these list of tokens stop words are removed which would give us the required set of words which are further put into a process called stemming.

Stemming is a normalization technique where the set of words obtained from the previous process is being normalised to convert it into a sequence to shorten its lookup**.**

➤ *Step 2*: *Conversion of text to speech*

Once the text is matched with the training data, the system has the ability to fetch the information that needs to be provided and responds accordingly.
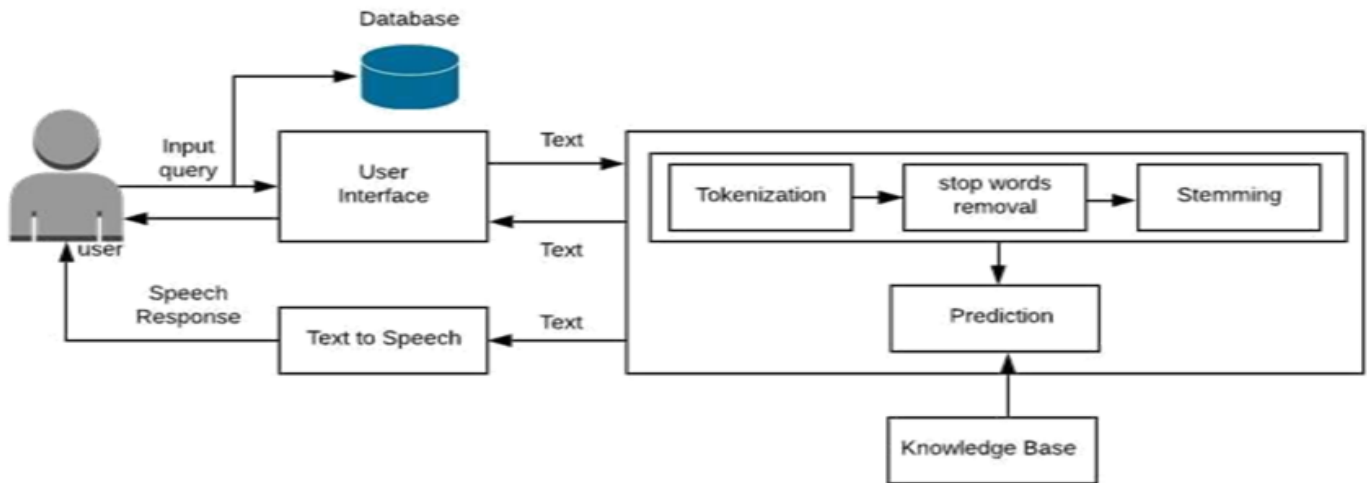

Fig 1:- System architecture

Pre-processing steps implemented in our system includes:

➤ *Tokenization*
In this process the sentence is split into its constituent words. During this step, characters like punctuation marks, white spaces, numbers are removed from the sentence.

➤ *Removal of stop words*
In this process words like "the", "on", "a", "all", "is" etc,. which do not add any meaning to the sentence and only increase the implementation time are stripped off so that the size of the input words are reduced and hence reducing complexity.

➤ *Stemming*
Stemming is the process wherein words are reduced to their root/base form. Our system uses Porter stemming algorithm. This algorithm removes inflectional endings from words.

After successful completion of the above said steps the remaining words are matched with the patterns already present in the knowledge base for retrieval of the most suitable response to the given query. The appropriate response is provided both in the form of text and speech to the user. In order to achieve this, we have made use of

IV. METHODOLOGY

In our system, the user will ask a query as a voice input in the user interface. Later the user input query is pre-processed to enable pattern matching.

Natural Language Toolkit (NLTK) There are two main modules in our system namely Admin and the Principal

➤ *Flowchart for admin:*

The Admin can login using the credentials: name and password that enables the admin to access his page. If the credentials are not matched, login is denied. The admin is then allowed to register principal details and also has the authority to add or delete principal details.
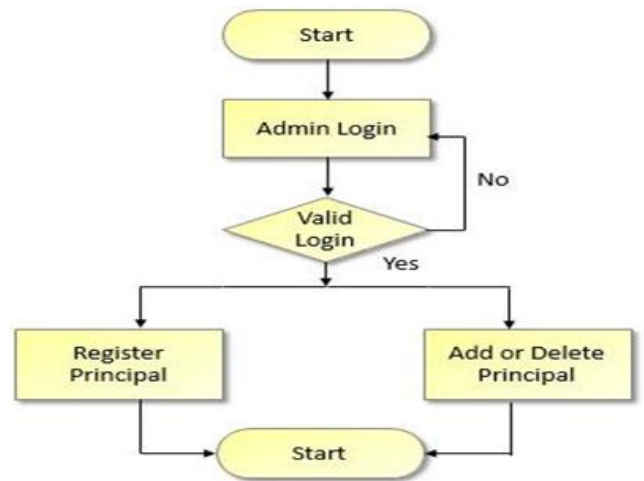

Fig 2

➤ *Flowchart for principal:*

The principal can login using name and password. If the credentials do not match, he/she is denied login. The principal can then open the database, update the database and view the queries.
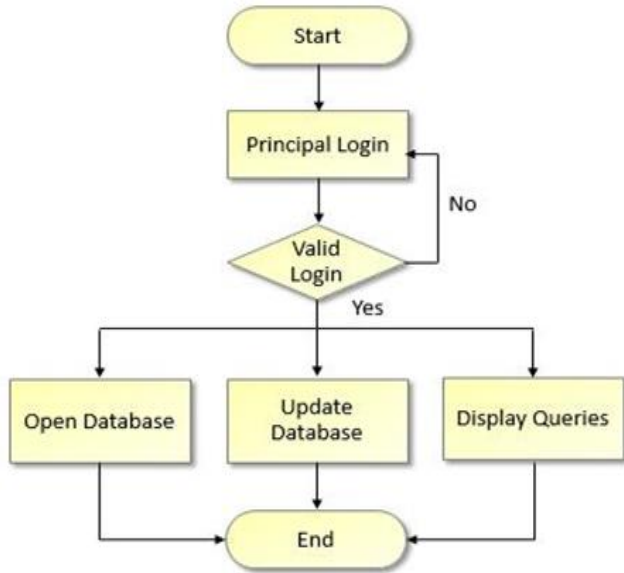
Fig 3

V. RESULTS

Fig 4

The Fig 4. depicts two logins present namely, admin login and principal login. The home page also includes three categories for the user to choose his query from. They are directions, general information and principal availability.
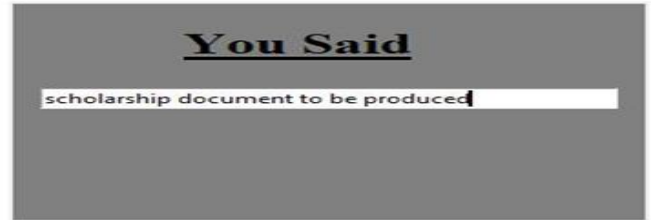
Fig 5

The Fig 5 depicts the page where the user can ask his query as a voice input. Later this query is converted into text for the user to view.

Fig 6

The Fig 6. illustrates the page where appropriate response to the query asked is provided both in the form of text and speech.
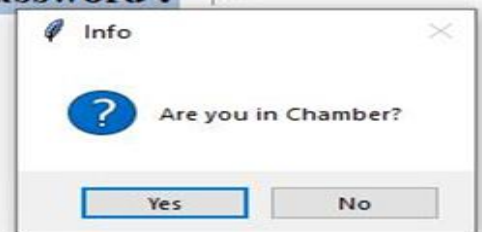
Fig 7

The Fig 7. depicts the page where once the principal logs in, there is a pop up asking if he is in the chamber or not. He can choose either yes or no which would result changes in the principal availability module.

Fig 8

The Fig 8. represents the operations that can be performed by the principal after his successful login. The operations include opening his database, updating the database and viewing the queries intented for him to respond later.

VI.   CONCLUSION

The system is implemented to give an automatic response to the user using machine learning. The user will ask his/ her queries and then the appropriate keyword from the given query is extracted to produce the suitable response. The future scope can be implemented by adding parallel processing which enables multiple users to access the system simultaneously. The system can also make use of artificial intelligence to fetch the information that is not present in the training data. As a result of this, every answer for the user's query will be generated either from static database or through online sources.

REFERENCES

[1]. A. R. Memon, B. S. Chowdhry, S. M. S. Shah, T. R. Memon and S. M. Z. A. Shah, "An electronic information desk system for information dissemination in educational institutions," *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2015, pp. 1275-1280.

[2]. S. Ghose and J.J. Barua, "Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor", 2013 International Conference on Informatics, Electronics & Vision (ICIEV), 2013

[3]. Aakash G Ratkal, GangadharAkalwadi, Vinay N Patil and Kavi Mahesh, (2016), "Farmer's Analytical Assistant", IEEE International Conference on Cloud Computing in Emerging Markets.

[4]. BayuSetiaji, Ferry WahyuWibowo, (2017), "Chatbot Using A Knowledge in Database - Human - to - Machine Conversation Modeling", 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS).

[5]. Naveen Kumar M, LingaChandar P C, Venkatesh Prasad A, Sumangali K, (2016), "Android Based Educational Chatbot for VisuallyImpairedPeople",IEEE International Conference on Computational Intelligence and Computing Research.

[6]. D. Jurafsky, I. D. Jurafsky, J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics and Speech Recognition, Upper Saddle River, NJ:Prentice Hall, pp. 988, 2008.

[7]. M. Maldonado, D. Alulema, D. Morocho and M. Proaño, "System for monitoring natural disasters using natural language processing in the social network Twitter," *2016*