

Chatbot for college related FAQs

¹Prof. Vishal R. Shinde, ² Miss. Anagha Bagul, ³ Mr. Amit Gupta, ⁴ Miss. Sneha Javeri

¹Asst.Professor, ^{2,3,4}UG Student, ^{1,2,3,4}Computer Engg. Dept. Shivajirao S.Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

¹ *mailme.vishalshinde@gmail.com*, ² *anaghabagul54@gmail.com*, ³ *amitg.2797@gmail.com* ,
⁴ *javerisneha1@gmail.com*

Abstract- A Chatbots is a conversational representative that act together with users using natural language. Traditionally to solve a query by a software program it use to involve search engine of filling outer form. A chatbot permits a user to merely ask a query in a same style that they would talk a human. Chatbots have turn out to be more widespread in business groups now as they reduce customer service cost and handle multiple users at a time[1]. Recent advancement in machine learning and artificial intelligence have significantly enhanced the correctness and efficiency of natural language processing (NLP), making chatbot a great option for many organizations. This paper grants the proposal of a chatbot, which offers proficient and precise answer for any query/FAQs related to a college by the help of machine learning and neural network using python. This system makes use of feedback-feedforward technique to improve the exactness of the chatbot. This chatbot can be operated for any college to answer FAQs to interested students in a great fashion.

Keywords- Chatbot, Corpus, Neural Network

I. INTRODUCTION

Artificial intelligence aims to create machines which would have human intelligence, one example model of this is a chatbot. Chatbot is a trending application which has been developed using the concept of Artificial Intelligence. The chatbot is more supportive when it arises to educational tenacity. Principally a chatbot is one computer program that is created to stimulate intellectual human language collaboration through text or speech and whose purpose is to participate in conversations or to emulate in proper chat message between a human user and a computer by means of natural language. Chatbot is a contrived material that is proposed to mimic a clever decision with human assistants through their regular language [1]. A chatbot is a service driven by instructions and sometimes AI that you cooperate via a chat interface. The facility could be any purposeful, administrative or entertaining[2]. Chatbots have the benefit that it can quiet simply be used in any business one has to merely train the bot by providing the true conversation structure and drift to alter its current arena or business. AIML, XML, Python and many more languages can be used for developing a chatbot. The proposed system would be developed using python. The system basically makes use of sequence to sequence/generative model to identify the keywords from the user entered query. This chatbot would be developed for college students to get basic details regarding college amenities, faculty, location & their facilities.

II. LITERATURE SURVEY

Paper 1: Chatbot for University Related FAQs

This paper has provided the scheme of a chatbot, that delivers an resourceful and precise answer for any query built on the dataset of FAQs by means of Artificial Intelligence Markup Language (AIML) and Latent Semantic Analysis (LSA)[1].

Paper 2: Real World Smart Chatbot for Customer Care using a Software as a Service (SaaS) Architecture.

This chatbot would analyse messages of the user to check whether it is solvable or not according to the predefined data. If yes then chatbot initializes the interaction using LUIS and cognitive services. This system is implemented over AWS public cloud[2].

Paper 3: An Internet Relay Chatbot Using AIML:

It is an android based chatbot. Provide education grounded chatbot for visually weakened persons. This system can be easily launched by using Google Voice Search. The above system makes use of two types of resources mediawiki API, AB library. This method also used pattern matching technique[3].

Paper 4: Chatbot Using a Knowledge in Database:Human-to-Machine Conversation Modeling

The mechanism has been rooted information to recognize the sentences and making a choice itself as reply to answer a question. From input sentence, it will be scored

to get the resemblance of sentences, the higher score obtained the more similar of reference sentences. The sentence similarity calculation here in this research is done using bigram which divides input sentence as two letters of input sentence[4].

Paper 5: Smart Answering Chatbot based on OCR and Overgenerating Transformations and Ranking

This paper presents an Android based Chat- Bot for hotel booking. This paper focuses on automating the process of communication by use of chat-bot and it also emphases on giving adapted results to the operator which makes the procedure for hotel booking convenient and user friendly for him. This system attempts to create a chat-bot using Artificial Intelligence Markup Language and using various algorithms such as Keyword Matching. This system has been implemented to integrate with any Hotel Management Android Application to ease the process of hotel booking[5].

Paper 6: An intelligent web-based voice chat bot

This paper grants a tactic of altering documents into data of Chatbot system that permits operators to make additional benefits of it by questioning and replying questions via the usage of electronic documents combined with simulate system. It is an united system for enrich

contents of documents from popular format. The working of system is on track from excerpting texts using Optical Character Recognition (OCR) from files, then produce queries by Overgenerating Alterations and Ranking algorithm, and finally let Chatbot response to the user’s question when it is matched with the String pattern[6].

III. EXISTING SYSTEM

Until now all the information related to college activities, notices, exam related information, admission and scholarship related issues were solved only with the help of college notice board, WhatsApp group related to college , by college website or by a personal visit to the college . These methods were inefficient and too much time consuming. Sometime getting the required information is a failure from these resources. These resources aren't so efficient to answer every small doubt of the students. The existing system could not help the non-college members to know about the college and the daily activities carried out in the college. Thus, the proposed system would help students by solving their FAQs via scheming a Chabot that would help students to perform necessary queries without physically going to the college. No such chatbot has been designed for a college level purpose, so their is a great scope and need to develop such a chatbot system for the same.

IV. COMPARATIVE STUDY

Table 4.1: Comparative Analysis

Sr. No	Paper Title	Author’s Name	Algorithm/ Language/ Technology used	Advantages	Disadvantages
1	Real World Smart Chatbot for Customer Care using a Software as a Service (SaaS) Architecture	Bhavika R. Ranoliya, Nidhi Raghuvanshi and Sanjay Singh	LUIS and Cognitive services	Easy work frame and result generation is also quite fast.	Unable to handle huge number of requests at a time.
2	Android Based Educational Chatbot for Visually Impaired People	Naveen Kumar M, Linga Chandar P C, Venkatesh Prasad A, Sumangali K	Pattern Matching Algorithm	Can be used for both visually impaired and a normal vision person. Also it provides data from Wikipedia.	Input to the system is voice so there is a possibility of invalid word recognition so it may present a faulty result.
3	Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling	Bayu Setiaji ; Ferry Wahyu Wibowo	Sentence similarity Using bigram	This application work is very simpler because the knowledge already programmed in advance.	The utilization of MySQL database in the chatbot is just restricted to store the learning.
4	Smart Answering Chatbot based on OCR and Overgenerating Transformations and Ranking	Ly Pichponreay, Chi-Hwan Choi, Jin-Hyuk Kim, Kyung-Hee Lee, Wan-Sup Cho	Optical Character Recognition (OCR) and Ranking algorithm	Efficient and accurate, performance	Time consuming issue, information constraint problem, typing issue.

V. PROBLEM STATEMENT

1) Until now a computerized, digital information portal has been made available only on an university level. In

order to obtain information on an institute or college basis one had to be physically present at the said institute which is tedious as well as time consuming.

2) In order to tackle this hurdle a digital dynamic and interactive Chatbot has been developed which will provide all the relevant information on an immediate or as needed basis.

3) The Chatbot will allow the user to procure information locally in a dynamic and efficient layout which will be easy to understand and maintain for the administrator from the user's query keywords would be fetched and matched with the one present in the knowledge base and then an appropriate response for the query would be sent. The user can give his/her feedback by rating a particular answer. These ratings would be useful for further learning of the machine to improve its accuracy. If the keyword is not matched then a default message would be delivered "Answer Not Found" and the chatbot system would ask the user for his/her opinion for the answer which would be useful for further analysis and learning. Thus the proposed system makes use of unsupervised learning based on neural networks. The proposed system would be developed for any college and would guide students.

VI. PROPOSED SYSTEM

Proposed system would present a college FAQs chatbot which will help to answer the queries of the users that are related to the college in an efficient manner. The system would be developed using python. The chatbot system would use different APIs to get information from the database/corpus by using NLKT(Natural Language Tool Kit) library data would be parsed and various operations like cleaning, stemming and lemmatization would be performed using Numpy Class. Thus all the input data would be cleaned and processed for training. After processing the data, the data will be fed to a model implemented using sequence to sequence modeling method and neural nets. After training the data the model is ready for answering the questions

There are 2 types of users for the system:

- 1) General Users
- 2) Admin User

The general user has to first log in to the system then he/she can input query regarding the college via an input texted mode. Then It would help and guide the new students about different amenities, laboratories, classrooms or office locations inside the campus by providing the users with an image map. All the other basic queries would be answered back in a texted format. At the conclusion of conversation the user can give his/her feedback of the chat session.

VII. ALGORITHM

The general idea of working of proposed system algorithm is given as follow:

- Step 1: Start
- Step 2: Enter Query
- Step 3: Search Query="College Related FAQ";

- Step 4: if keyword = Keyword in corpus return(response) goto step 6
- else
- return ("Answer Not Found") goto step 5
- Step 5: User suggests an answer for query which was not found.
- Step 6: Give Rating for the query goto step 7 or step 2
- Step 7: Exit.

VIII. MATHEMATICAL MODEL

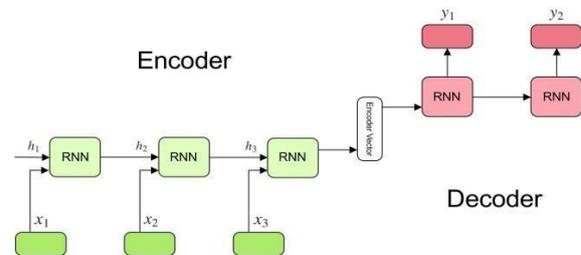


Fig 1: Sequence to Sequence Typical working

The model mainly comprises of 3 measures: encoder, intermediate (encoder) vector and decoder.

Encoder admits a single portion of the input sequence, gathers information for that portion and sends it forward. In question- answering problem, the input stream is a collection of words of the question. Each word is represented as x_i where i denoted the word's order

The hidden states h_i are computed using the formula:

$$h_t = f(W^{(hh)}h_{t-1} + W^{(hx)}x_t)$$

Encoder Vector is the last hidden state formed from the encoder portion. This vector aims to summarize the data for all input portions in order to support the decoder make precise predictions. It turns as the early hidden state of decoder portion of the model.

Decoder: A stack of numerous recurring components where each calculates an output y_t at a time step t . Individually recurrent unit takes a hidden state from the earlier unit and yields an yield as well as its personal hidden state. In question problem, the output arrangement is a group of words from the response. Each word is represented as y_i where i is the order of that word. Any hidden state h_i is computed using the formula:

$$h_t = f(W^{(hh)}h_{t-1})$$

The output y_t at time step t is calculated using the formula:

$$y_t = \text{softmax}(W^S h_t)$$

Softmax is used to make a likelihood vector that aids us determine the final output

Sequence to Sequence Method working code:

```
# In[2]:
def build_dataset(words, n_words):
```

```
count = [['GO', 0], ['PAD', 1], ['EOS', 2],
['UNK', 3]]
```

```
count.extend(collections.Counter(words).mo
st_common(n_words - 1))
dictionary = dict() for
word, _ in count:
    dictionary[word] = len(dictionary) data =
list()
unk_count = 0
for word in words:
    index = dictionary.get(word, 0) if index
== 0:
        unk_count += 1
        data.append(index)
    count[0][1] = unk_count
    reversed_dictionary =
dict(zip(dictionary.values(),
dictionary.keys()))
return data, count, dictionary,
reversed_dictionary
```

```
# In[3]:
with open('from.txt', 'r') as fopen:
    text_from = fopen.read().lower().split("\n") with
open('to.txt', 'r') as fopen:
    text_to = fopen.read().lower().split("\n") print("len
from: %d, len to:
%d"%(len(text_from), len(text_to))) concat_from =
''.join(text_from).split() vocabulary_size_from =
len(list(set(concat_from)))
ta_from, count_from, dictionary_from,
rev_dictionary_from = build_dataset(concat_from,
vocabulary_size_from)
concat_to = ''.join(text_to).split()
vocabulary_size_to =
len(list(set(concat_to)))
data_to, count_to, dictionary_to, rev_dictionary_to =
build_dataset(concat_to, vocabulary_size_to)
GO = dictionary_from['GO'] PAD =
dictionary_from['PAD'] EOS =
dictionary_from['EOS'] UNK =
dictionary_from['UNK']
```

IX. SYSTEM ARCHITECTURE

Description: The system architecture consist of 4 main parts:

1. Users.
2. Admin.
3. Chatbot System.
4. Corpus

A User can enter a query related to the college in text format. The query is then sent to the Corpus, the corpus consist of all the information related to the college in an encoded format , via use of sequence-to- sequence model in the Chatbot System. The Result for Query is searched

in college data present inside the Corpus. This process is done by keyword matching Technique. If Keyword matching to the query is found then the info related to the query is sent as a response in text format. But if Template matching to Query is not found then a default message is displayed. I User thinks that the result of the query is irrelevant then he/she can give Rating to the answer present on the GUI. The data is called actionable since the user can suggest the Admin that the response given was invalid that was his opinion, this is done by user by giving less rating on the rating bar. The analysis of these responses are noted using feedback- feedforward system and an improvement is done by the Chatbot using learning. Thus the system undergoes changes by the process of unsupervised learning.

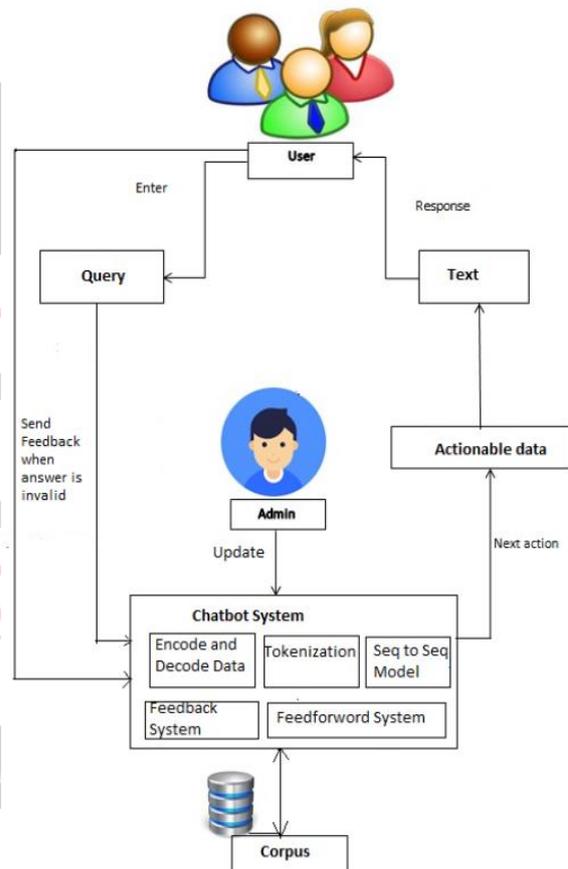


Fig.2: System Architecture

X. ADVANTAGES

- 1) The Chatbot is a timer saver for all the users i.e. students, parents, teaching and non-teaching staff
- 2) The chatbot helps the students to be updated with all the college events and activities.
- 3) The chatbot helps the enquirer a lot since they need not go personally to the college office for the enquiry.
- 4) The chatbot helps the new students a lot by providing them the maps that guide them to different laboratories and amenities.

