COLLEGE ENQUIRY CHATBOT

Submitted By
B.Venkata Pavani R170431
M.Anitha R170819

Under the Guidance of

Ch.Ratna Kumari

Department of Computer Science and Engineering

Rajiv Gandhi University of Knowledge Technologies (RGUKT),
R.K.Valley , Kadapa , Andhra pradesh.

as a part of
Partial fulfillment of the degree of Bachelor of
Technology in Computer Science and Engineering

Date : 19-09-2022

CERTIFICATE OF PROJECT COMPLETION

This is to certify that the report entitled **"College Enquiry Chatbot"** submitted by **B.V.Pavani** bearing ID.No. **R170431**, **M.Anitha** bearing ID.No. **R170819** and in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out by them under my supervision and guidence.

The report has not been submitted previously in part or in full to this or any other University or Institution for the award of any degree or diploma.

Project Guide,

Ch.Ratna Kumari,

Assistant Professor,

Dept of CSE,

RGUKT RK.Valley,

Head of the Department ,

P.Harinadha,

Assistant Professor,

Dept of CSE,

RGUKT RK.Valley,

Declaration

We, hereby declare that this report entitled **"College Enquiry Chatbot"** submitted by us under the guidance and supervision of **Ch.Ratna Kumari** is a bonafied work. We also declare that it has not been submitted previously in part or in full to this university or other university or institution for the award of any degree or diploma.

We will be solenly responsible if any kind of plagiarism is found.

Date:19-09-2022 R170431,B.V.Pavani
Place:RK Valley R170819,M.Anitha

Acknowledgement

We would like to express our sincere gratitude to **Ch.Ratna Kumari**, Our project guide for valuable suggestions and keen interest throughout the period of project work.

We are grateful to **P.Harinadha, HOD OF CSE** for providing excellent computing facilities and a congenial atmosphere for progressing with our project.

At the outset, I would like to thank **Rajiv Gandhi University of Knowledge Technologies,RK Valley** for providing all the necessary resources for the successful completion of our project work.

We express our thanks to all those who contributed for the successfull completion of our project work.

With gratitude,

B.V.Pavani R170431
M.Anitha R170819

ABSTRACT

In this work, the proposal is made to implement College Enquiry Chatbot using NLP(Natural Language Processing) concept. Natural Language Processing (NLP) is a component of AI in the field of linguistics that deals with interpretation and manipulation of human speech or text using software. It enables the computer to understand the natural way of human communication by combining machine learning, deep learning and statistical models.Smart assistants and Email Filters are some of the applications of NLP. In this work,we use few modules for analyzing, tokenizing and for lemmatization of the given text. The model is Sequential model built using NLP based on bagOfWords algorithm. The implementation of NLP Chatbot is demonstrated and analyzed to study the various feautures of Chatbots. This chatbot is implemented using python language in google colaboratory.

Table of Contents

1.Introduction.....**7**

1.1.Introduction to the project
1.2.Importance of the project

2.Preliminaries.....**8-11**

2.1.Importing Python Libraries
2.2.Creating JSON file
2.3.Preprocessing
2.4.Designing Neural Network

3.Proposed Model.....**12-13**

3.1.Note on JSON dataset
3.2.About Sequential Model

4.Block Diagram for NLPChatbot.....**13**

5.Flowchart for NLPChatbot.....**14**

6.Implementation of Model.....**15-18**

7.Result.....**19-20**

8.Testing.....**21-22**

9.Conclusion.....**23**

10.Future Scope.....**24**

11.References.....**25**

INTRODUCTION

1.1 Introduction to project

Chatbot was a computer application which may speak to human beings naturally, the way we interact with one another . It can replace a person's for several tasks of answering queries. A chatbot is an agent that interacts with users using simple language. Several applications of chatbots like Customer Service, call centers etc. uses AI terminology to talk with user. one among the prime goals of chatbots is to resemble an intelligent human and make it difficult for the receiver of the conversation to know the important working along side various architecture and capabilities for his or her usage has widely broadened.

In this work,we created Chatbot for Rajiv Gandhi University Of Knowledge Technologies, Rk Valley college. Here, we use Sequential model and design a neural network. Here we use Natural Language Processing(NLP) modules for tokenization and lemmatization of the text.

We use numpy module to represent the text in number array form.

1.2 Importance of the Project

The College Enquiry Chatbot has the capacity to make friendly conversations; respond the course and faculty details; give the link for the academic calendar; answer the frequently asked questions; calculate provide the details about the fees based on the student’s input; and give the timings, address, contacts, and events information of the departments like Union, Library.

2.PRELIMINARIES

2.1.Importing Python Libraries

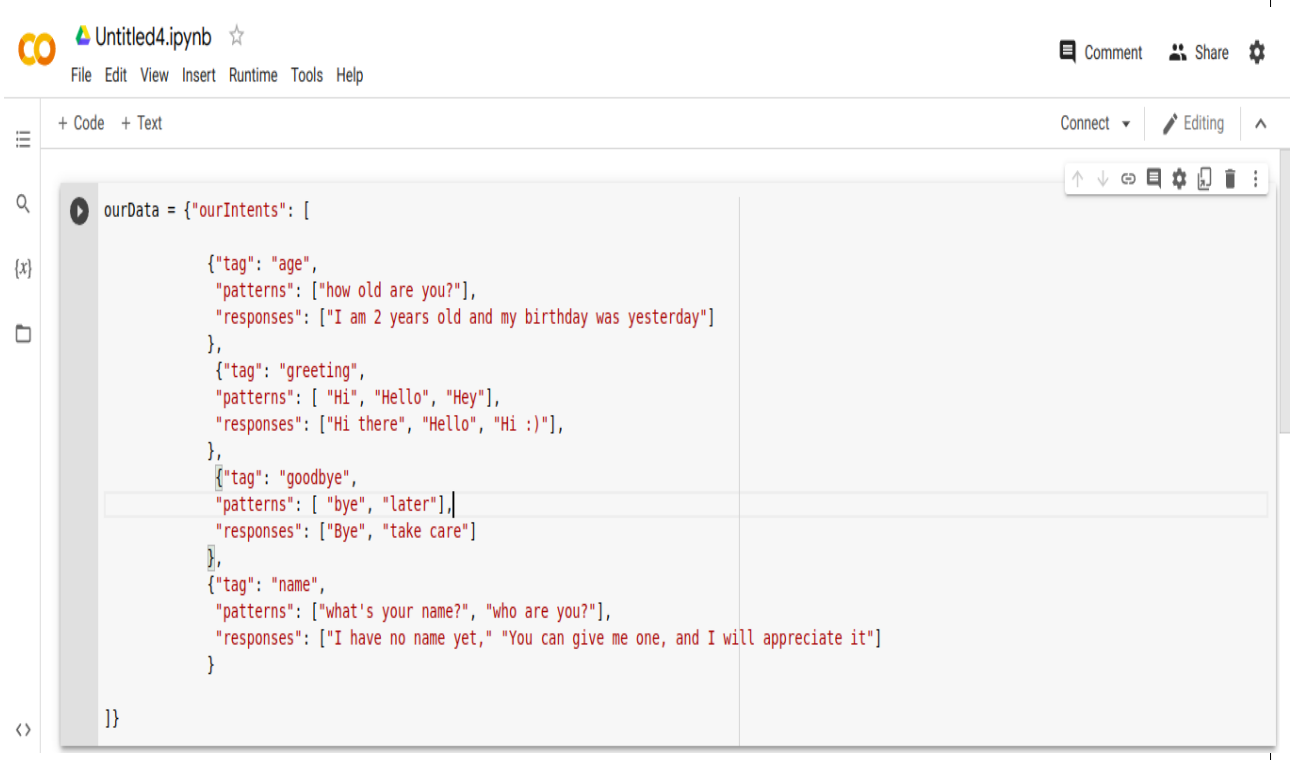
Imports are critical for successfully organizing your Python code. Correctly importing code will increase your productivity by allowing you to reuse code while also maintaining the maintainability of your projects.

The necessary libraries include:

- JSON: It is possible to utilize it to work with JSON data. A JSON file stores data in key-value pairs and arrays; the software it was made for then accesses the data. JSON allows developers to store various data types as human-readable code, with the keys serving as names and the values containing related data.
- String: Provides access to several potentially valuable constants. It's a built-in module and we have to import it before using any of its constants and classes.
- Random: For various distributions, this module implements pseudo-random number generators. Python Random module is an in-built module of Python which is used to generate random numbers. These are pseudo-random numbers means these are not truly random. This module can be used to perform random actions such as generating random numbers, print random a value for a list or string, etc.
- WordNetLemmatizer: It can lemmatize.Lemmatization is a text normalization technique used in Natural Language Processing (NLP), that switches any kind of a word to its base root mode. Lemmatization is responsible for grouping different inflected forms of words into the root form, having the same meaning.
- Tensorflow: A multidimensional array of elements is represented by this symbol. TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow.
- Sequential: Sequential groups a linear stack of layers into a tf.keras.Model. The core idea of Sequential API is simply arranging the Keras layers in a sequential order and so, it is called Sequential API. Most of the ANN also has layers in sequential order and the data flows from one layer to another layer in the given order until the data finally reaches the output layer.

2.2.Creating a JSON file

This step will create an intents JSON file that lists all the possible outcomes of user interactions with our chatbot. We first need a set of tags that users can use to categorize their queries. These tags include name , age , and many others. Every new tag would require a unique pattern.



2.3.Preprocessing

In this section, vocabulary of all the terms used in the patterns, list of tag classes, list of all the patterns in the intents file, and all the related tags for each pattern will be created before creating our training data.

Each intent is tokenized into words and the patterns and their associated tags are added to their respective lists.

2.4.Designing Neural Network

We design a neural network from sequential model for the chatbot, because neural networks can only understand numerical values, we must first process our data so that a neural network can understand what we are doing.

After converting our data to a numerical representation, we can now design a neural network model which we will feed our training data. The model will select an appropriate response from the tag associated with a given feature.

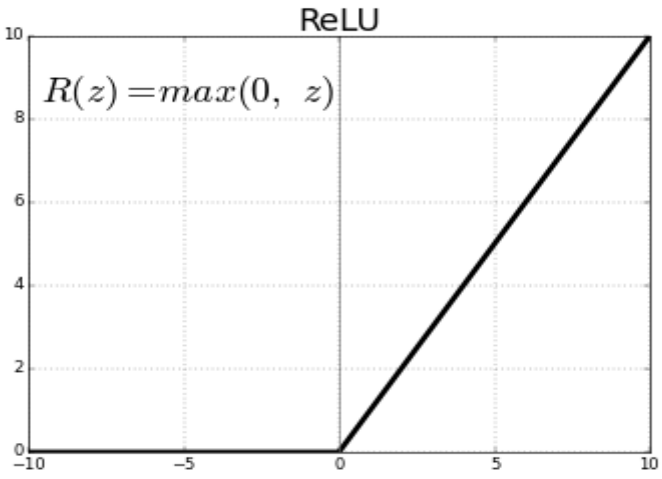
Here we use Sequential model for designing neural network. Here we create three layers which contains two middle layers and one output layer. The two inner layers use Relu activation Function and the output layer uses softmax activation function.

Here we use mainly two activation functions :-

- Relu activation functions
- softmax activation function

Relu activation function :-

The rectified linear activation function or ReLU for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero.The usage of ReLU helps to prevent the exponential growth in the computation required to operate the neural network.The main reason why ReLu is used is because it is simple, fast, and empirically it seems to work well.ReLU is not linear. The simple answer is that ReLU 's output is not a straight line, it bends at the x-axis. The more interesting point is what's the consequence of this non-linearity. In simple terms, linear functions allow you to dissect the feature plane using a straight line.



Three reasons I choose ReLU as an Activation Function

- First it's Non-Linear(although it's acts like a linear function for x > 0)
- ReLU is cheap to compute. Since it's simple math, model takes less time to run
- ReLU induces sparsity by setting a min value of 0

Softmax activation function:-

Softmax is an activation function that scales numbers/logits into probabilities. The output of a Softmax is a vector (say v) with probabilities of each possible outcome.

Formula:-

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

σ = softmax

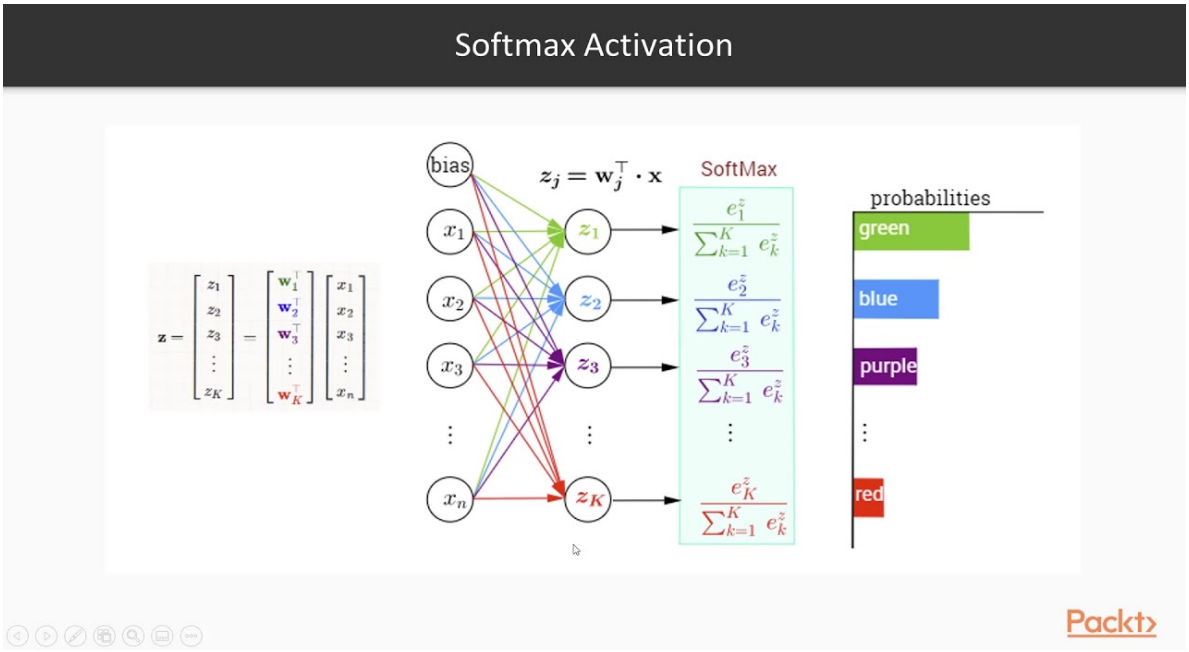
$\vec{z}$ = input vector

e^{z_i} = standard exponential function for input vector

K = number of classes in the multi-class classifier

e^{z_j} = standard exponential function for output vector

e^{z_j} = standard exponential function for output vector



The main advantage of using Softmax is the output probabilities range. The range will 0 to 1, and the sum of all the probabilities will be equal to one. If the softmax function used for multi-classification model it returns the probabilities of each class and the target class will have the high probability.

3.Proposed Model

3.1.About JSON Dataset

In this NLP Chatbot, we use JSON file Dataset which is created according to the facilities provided by the Rajiv Gandhi University Of Knowledge Technologies, Rk Valley College. This dataset is created by using tag, patterns and corresponding responses for each tag. The example for the JSON dataset is given above in creating a JSON file section.

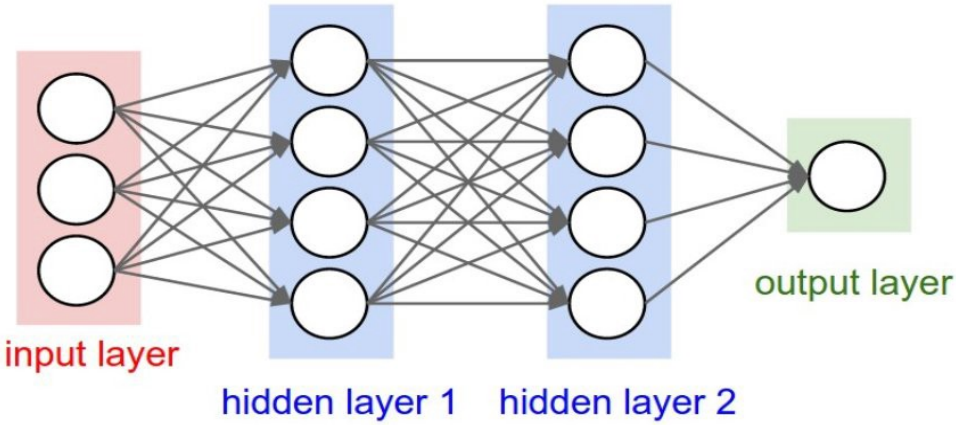
We create an intents array in which each element comtains tag,pattern and corresponding responses. The output of the chatbot is produced by preprocessing this JSON Dataset.

3.2. About Sequential Model

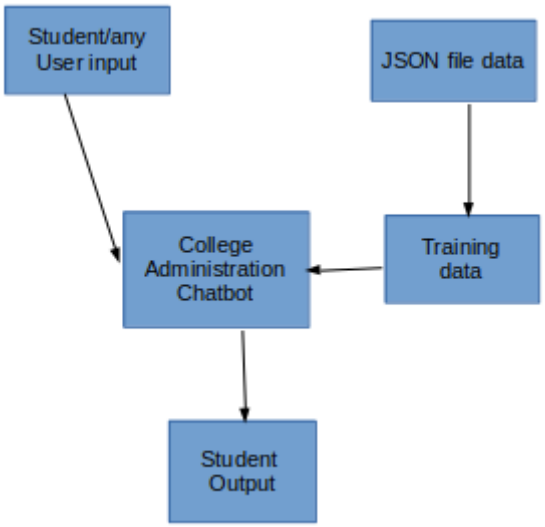
Machine learning models that input or output data sequences are known as sequence models. Text streams, audio clips, video clips, time-series data, and other types of sequential data are examples of sequential data. Recurrent Neural Networks (RNNs) are a well-known method in sequence models.

A Sequential model is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor. A Sequential model is not appropriate when: Your model has multiple inputs or multiple outputs. Any of your layers has multiple inputs or multiple outputs.

Creating a Sequential Model using Tensorflow:-

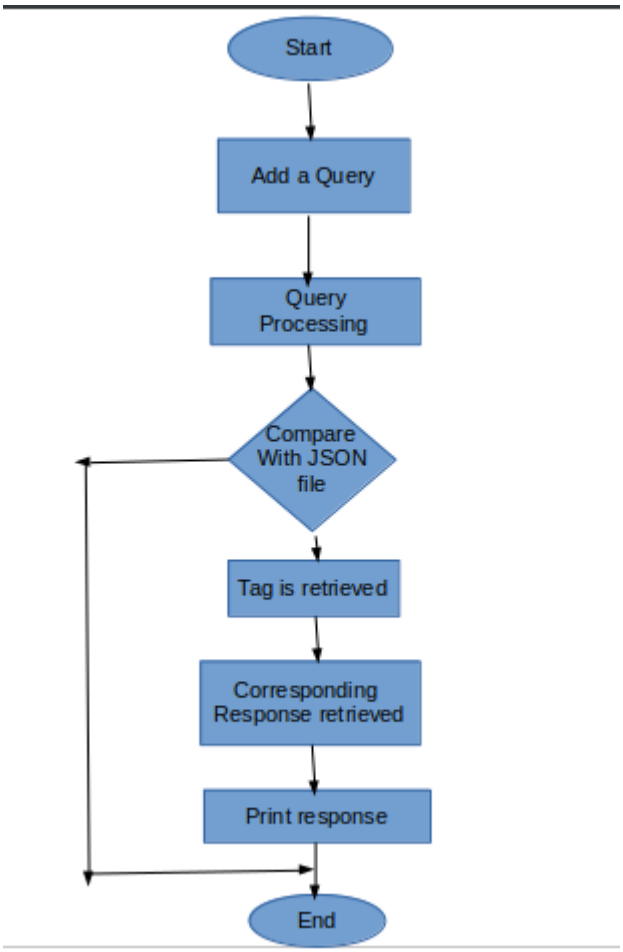


4. Block Diagram for NLP Chatbot



From the block diagram it is clear that, if student or any other user provides input to the Chatbot then that input will be stored in the database , the answer related to the query will be displayed to the user.the entire background process will be done by the admin.

5.Flowchart for NLP Chatbot



The above Flow Chart describes the entire process of the system, if the user query is not found in the database then we will collect the details from the user and one of the person from the college will contact the user personally. The answered questions will be stored in the database and the corresponding answers for those queries, will be updated by the Admin.

6.Implementation of the Model

The necessary libraries include:

- JSON: It is possible to utilize it to work with JSON data.
- String: Provides access to several potentially valuable constants.
- Random: For various distributions, this module implements pseudo-random number generators.
- WordNetLemmatizer: It can lemmatize.
- Tensorflow: A multidimensional array of elements is represented by this symbol.
- Sequential: Sequential groups a linear stack of layers into a tf.keras.Model.

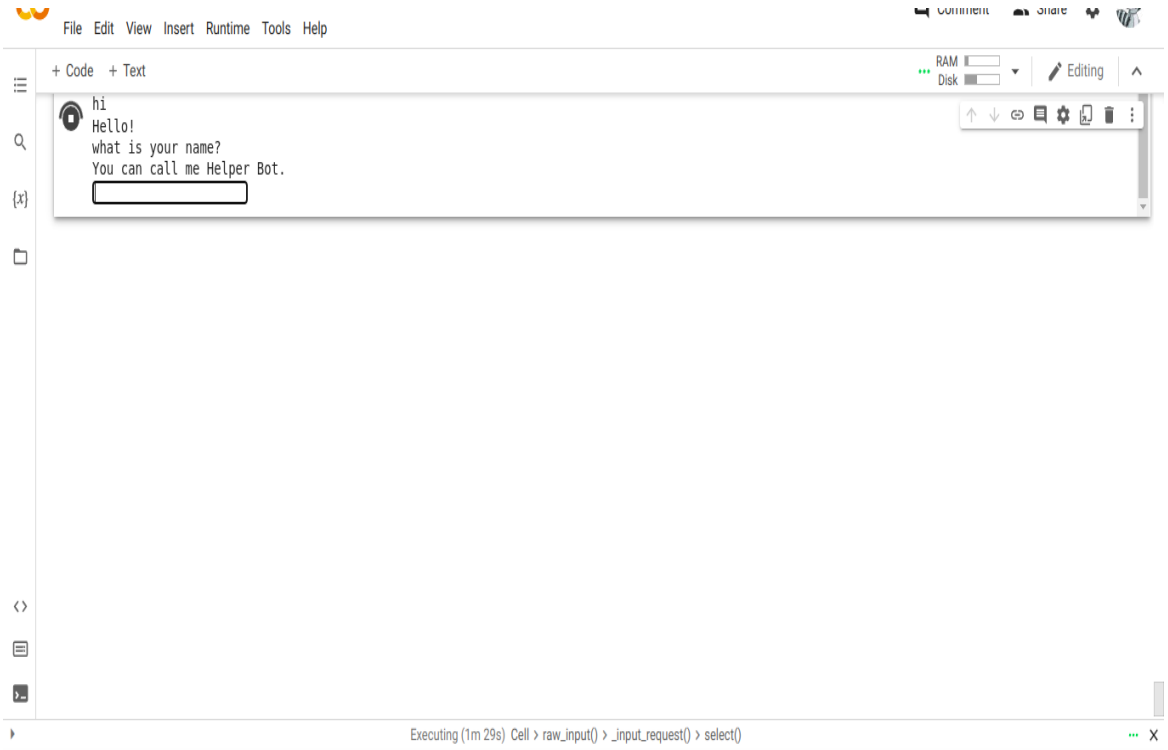
```
import json
import string
import random
import nltk
import numpy as num
from nltk.stem import WordNetLemmatizer
import tensorflow as tensorF
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense, Dropout

nltk.download('omw-1.4')
nltk.download("punkt")
nltk.download("wordnet")

data= {
        "intents": [
        {
                "tag": "greeting",
                "patterns": [
                                "Hi",
                                "How are you",
                                "Is anyone there?",
```

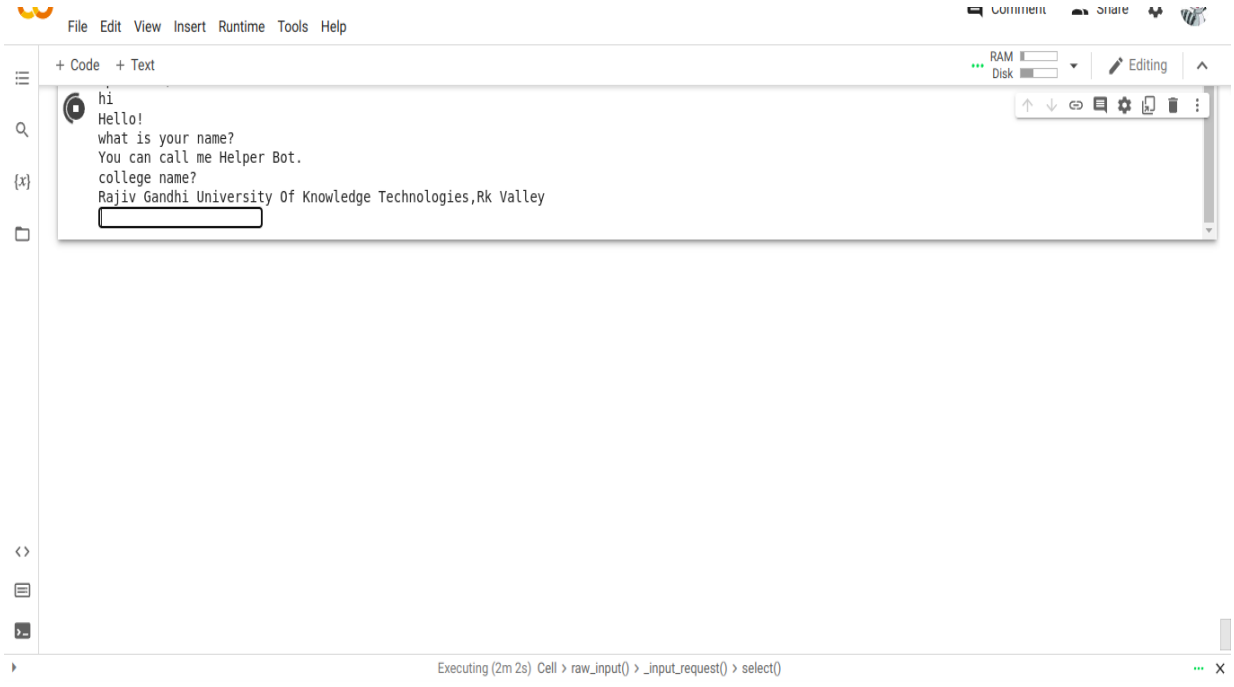
```
                    "Whats up",
                    "how are ya",
                    "heyy",
                    "whatsup"
                    ],
            "responses": [
                    "Hello!",
                    "Good to see you again!",
                    "Hi there, how can I help?"
            ],
            "context_set": ""
        },
        ........
    ]}
lm = WordNetLemmatizer()
ourClasses = []
newWords = []
documentX = []
documentY = []
for intent in data["intents"]:
        for pattern in intent["patterns"]:
                ournewTkns = nltk.word_tokenize(pattern)
                newWords.extend(ournewTkns)
                documentX.append(pattern)
                documentY.append(intent["tag"])
        if intent["tag"] not in ourClasses:
                ourClasses.append(intent["tag"])
newWords = [lm.lemmatize(word.lower()) for word in newWords if word not in
string.punctuation]
newWords = sorted(set(newWords))
ourClasses = sorted(set(ourClasses))
trainingData = []
```

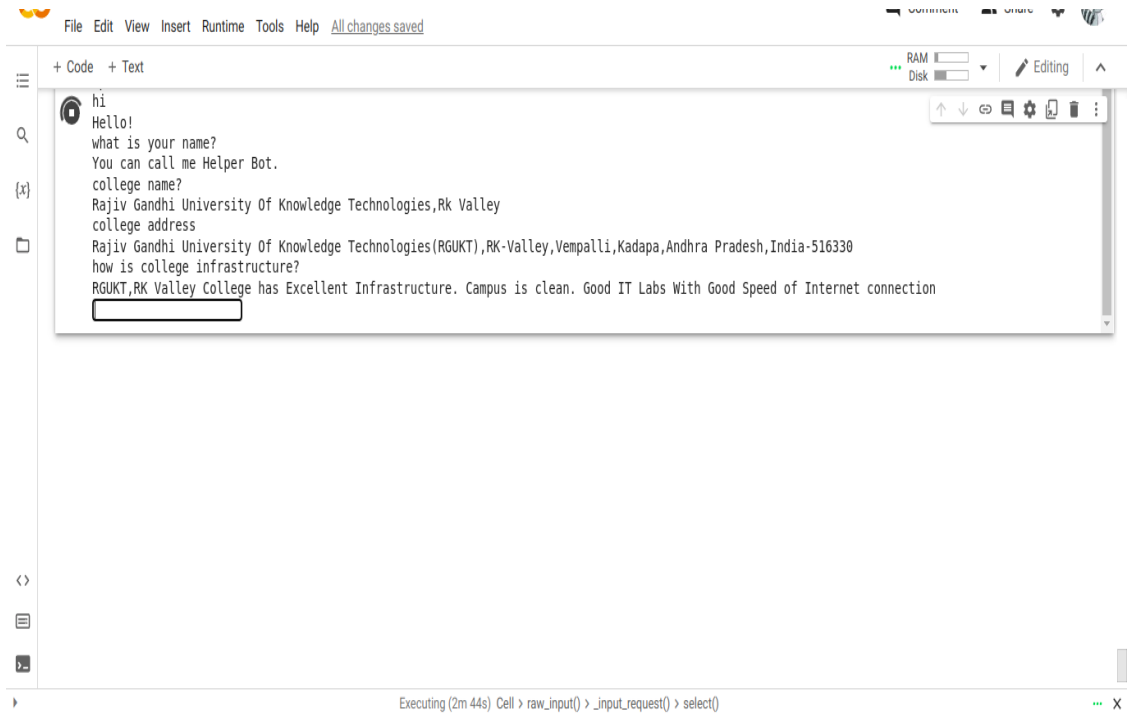
```
outEmpty = [0] * len(ourClasses)
for idx, doc in enumerate(documentX):
        bagOfwords = []
        text = lm.lemmatize(doc.lower())
        for word in newWords:
                bagOfwords.append(1) if word in text else bagOfwords.append(0)
        outputRow = list(outEmpty)
        outputRow[ourClasses.index(documentY[idx])] = 1
        trainingData.append([bagOfwords, outputRow])

random.shuffle(trainingData)
trainingData = num.array(trainingData, dtype=object)

x = num.array(list(trainingData[:, 0]))
y = num.array(list(trainingData[:, 1]))

iShape = (len(x[0]),)
oShape = len(y[0])

ourNewModel = Sequential()
ourNewModel.add(Dense(128, input_shape=iShape, activation="relu"))
ourNewModel.add(Dropout(0.5))
ourNewModel.add(Dense(64, activation="relu"))
ourNewModel.add(Dropout(0.3))
ourNewModel.add(Dense(oShape, activation = "softmax"))

md = tensorF.keras.optimizers.Adam(learning_rate=0.01, decay=1e-6)

ourNewModel.compile(loss='categorical_crossentropy',
optimizer=md,
metrics=["accuracy"])
print(ourNewModel.summary())
ourNewModel.fit(x, y, epochs=200, verbose=20)
```

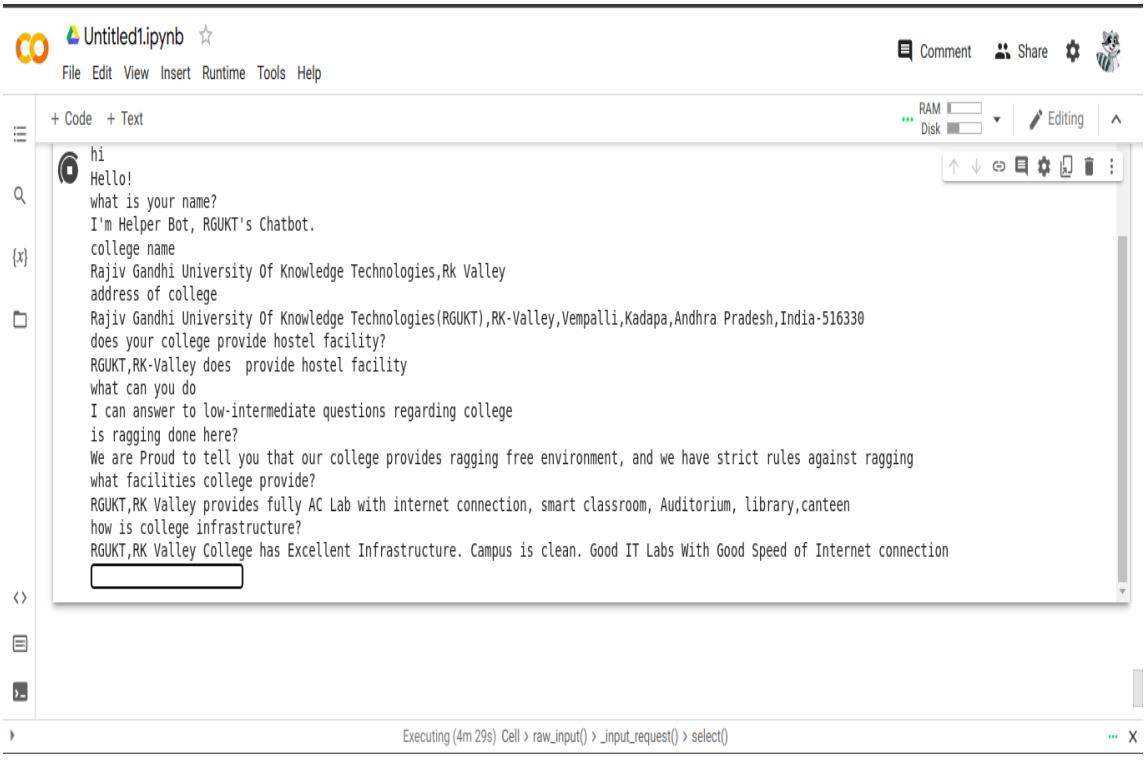
```
def ourText(text):
      newtkns = nltk.word_tokenize(text)
      newtkns = [lm.lemmatize(word) for word in newtkns]
      return newtkns
def wordBag(text, vocab):
      newtkns = ourText(text)
      bagOwords = [0] * len(vocab)
      for w in newtkns:
            for idx, word in enumerate(vocab):
                  if word == w:
                        bagOwords[idx] = 1
      return num.array(bagOwords)
def Pclass(text, vocab, labels):
      bagOwords = wordBag(text, vocab)
      ourResult = ourNewModel.predict(num.array([bagOwords]))[0]
      newThresh = 0.2
      yp = [[idx, res] for idx, res in enumerate(ourResult) if res >
      newThresh]

      yp.sort(key=lambda x: x[1], reverse=True)
      newList = []
      for r in yp:
            newList.append(labels[r[0]])
      return newList
def getRes(firstlist, fJson):
      tag = firstlist[0]
      listOfIntents = fJson["intents"]
      for i in listOfIntents:
            if i["tag"] == tag:
                  ourResult = random.choice(i["responses"])
                  break
      return ourResult
while True:
      newMessage = input("")
      intents = Pclass(newMessage, newWords, ourClasses)
      ourResult = getRes(intents, data)
      print(ourResult)
```

7.Result


File Edit View Insert Runtime Tools Help
+ Code + Text
RAM
Disk
Editing
hi
Hello!
what is your name?
You can call me Helper Bot.
Executing (1m 29s) Cell › raw_input() › _input_request() › select()


File Edit View Insert Runtime Tools Help
+ Code + Text
RAM
Disk
Editing
hi
Hello!
what is your name?
You can call me Helper Bot.
college name?
Rajiv Gandhi University Of Knowledge Technologies,Rk Valley
Executing (2m 2s) Cell › raw_input() › _input_request() › select()


File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
RAM
Disk
Editing
hi
Hello!
what is your name?
You can call me Helper Bot.
college name?
Rajiv Gandhi University Of Knowledge Technologies,Rk Valley
college address
Rajiv Gandhi University Of Knowledge Technologies(RGUKT),RK-Valley,Vempalli,Kadapa,Andhra Pradesh,India-516330
how is college infrastructure?
RGUKT,RK Valley College has Excellent Infrastructure. Campus is clean. Good IT Labs With Good Speed of Internet connection
Executing (2m 44s) Cell › raw_input() › _input_request() › select()


Untitled1.ipynb
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Editing
hi
Hello!
what is your name?
I'm Helper Bot, RGUKT's Chatbot.
college name
Rajiv Gandhi University Of Knowledge Technologies,Rk Valley
address of college
Rajiv Gandhi University Of Knowledge Technologies(RGUKT),RK-Valley,Vempalli,Kadapa,Andhra Pradesh,India-516330
does your college provide hostel facility?
RGUKT,RK-Valley does provide hostel facility
what can you do
I can answer to low-intermediate questions regarding college
is ragging done here?
We are Proud to tell you that our college provides ragging free environment, and we have strict rules against ragging
what facilities college provide?
RGUKT,RK Valley provides fully AC Lab with internet connection, smart classroom, Auditorium, library,canteen
how is college infrastructure?
RGUKT,RK Valley College has Excellent Infrastructure. Campus is clean. Good IT Labs With Good Speed of Internet connection
Executing (4m 29s) Cell › raw_input() › _input_request() › select()

8.Testing

After all phase have been perfectly done, the system will be implemented to the server and the system can be used.

System Testing :-

The goal of the system testing process was to determine all faults in our project. The program was subjected to a set of test inputs and many explanations were made and based on these

explanations it will be decided whether the program behaves as expected or not. Our Project went through two levels of testing :-

1. Unit testing

2 .Integration testing

Unit Testing :-

Unit testing is commenced when a unit has been created and effectively reviewed. In order to test a single module we need to provide a complete environment i.e. besides the section we would require The procedures belonging to other units that the unit under test calls Non local data structures that module accesses. A procedure to call the functions of the unit under test with appropriate parameters.

1. Testing for the chatbot :-

Testing chatbot-Here we test whether the chatbot is giving the corresponding answers to the given queries or not by checking the resulted answers with the given JSON file.

Integration Testing :-

In the Integration testing we test various combination of the project module by providing the

input.The primary objective is to test the module interfaces in order to confirm that no errors are occurring when one module invokes the other module.

Test Case Report:-

S.No.	Test Cases	Input	Expected Result	Actual Result	Status	Remarks
1	Enter valid text “Hi”	“Hi”	Print message like replying to “Hi”	Display “Hello!”	Pass	No remarks
2	Accepting next messsage or not	“what is your name?”	Next entry should be taken	Next entry is taken	Pass	No remarks
3	Enter invalid text “00000”	“0000”	It should diplay “I don’t Understand ”	Displays “I don’t understa nd what you are saying”	Pass	No remarks
4	Pattern matching working or not	Any text	It should display response of the correspond ing tag it belong	It should display response to the correspo nding tag.	Pass	No remarks
5	Importing libraries working or not	Import random	It should import random module	It should import random module	Pass	No remarks

9.Conclusion

The proposed chatbot is successfully completed with reasonable accuracy. To conclude, College Enquiry Chatbot is helpful in guiding students with correct and most up to date sources of information. It is advantageous for international applicants for queries such as fee payment and academic matters. Students can get the information at their fingertips rather than visiting college office. It improves efficiency by taking over tasks for which humans are not essential.

10.Future Scope

To improve the current functionalities of College Enquiry Chatbot, in the future, the scope of the chatbot can be increased by inserting data for all the departments, training the bot with varied data, testing it on live website, and based on that feedback inserting more training data to the bot. Some of the new features which can be added to the bot are speech recognition feature through which students can ask their queries verbally and get the answers from the bot, integration with multiple channels such as phone call, SMS, and various social media platforms like Skype, Facebook and Twitter.

11.References

https://www.google.com
https://en.wikipedia.org
https://www.youtube.com