Chat-Bot

C++

Huzaifa Irfan (285501)
ME-11 C
SMME, NUST
Islamabad, Pakistan
huzaifairfan2001@gmail.com

Hassan Ahmad (289686)
ME-11 C
SMME, NUST
Islamabad, Pakistan

Hassaan Ahmad (284811)
ME-11 C
SMME, NUST
Islamabad, Pakistan

M. Abu Bakr Nadeem (291593)
ME-11 C
SMME, NUST
Islamabad, Pakistan

Abu Huraira Mukhtar (284624)
ME-11 C
SMME, NUST
Islamabad, Pakistan

***Abstract*—Chat-Bot is a software that uses Artificial Intelligence to chat with the humans. These type of softwares have applications in various fields such as telecommunication, e-commerce, customer-care etc.**

***Index Terms*—Python, C++, NLTK, JSON, AI**

I. Introduction

This document is a model for Python Based Chat-Bot Project under course CS-114 that answers queries related to c++ and can be extended to what ever queries database you want whether you want for e-commerce or customer-care services. This app is opensource and is available on the github repository. `https://github.com/HuzaifaIrfan/Chat-Bot-Cpp`

A. Classes

This program uses object oriented model for programming that includes classes, inheritence and polymorphism etc. By using classes, it became easier for an app developer to create class objects for different user at a same time [with different integrated database] and process between queries.

The Class Architecture of the program is shown in the figure Fig.1

B. Database

The Databases of this app is stored as JSON (JavaScript Object Notation) format. We have used JSON as it is simpler to use and create and has dynamic schemas. It can store nested objects and arrays. and store them as string in .json file extension. We have stored tokens, queries and responses all in a single file and seperate it using keys of JSON object.

The similar meaning tokens are stored in the Synonyms List and the queried tokens and their responses are stored in the Queries List as shown Fig. 2

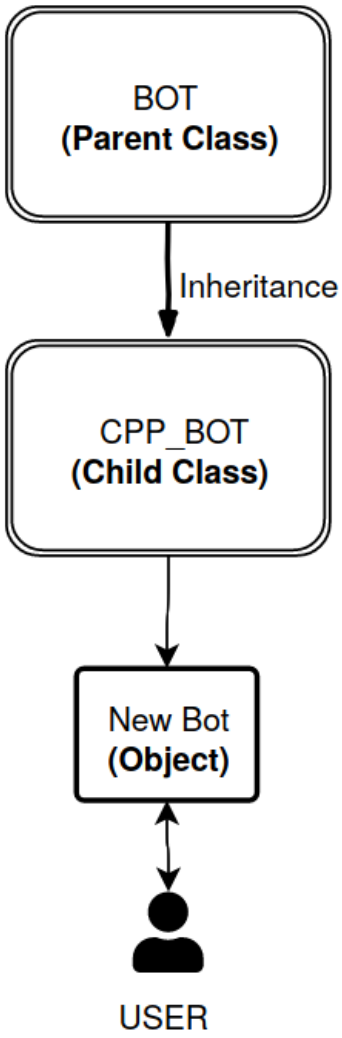

Fig. 1. Classes and Objects

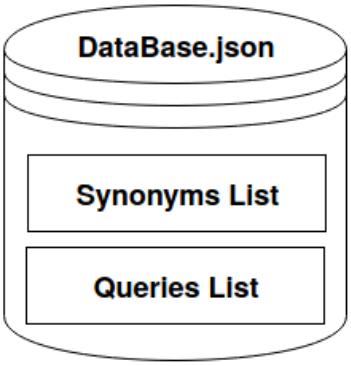

Fig. 2. DataBase Structure

C. Why use Python for development

- Python is high level programming language
- It is beginner friendly and easier to read and use.
- It supports multi-programming paradigms such functional or object oriented.
- It uses dynamic data types and memory management systems which allows users to focus on the algorithm.
- Python as an interpreted language supports and runs on many popular operating system without altering the code and re-compilations needed.
- PIP (Python Package Index) is preinstalled with python which makes it easier to install the external libraries.
- It is very fast growing language in the field of scientific computing, AI (Artifical Intelligence), Machine Learning and many other real world applications.

Fig. 3. Why Python

D. Limitations

- The app cannot process between two or more queries at same time.
- This doesn't uses Machine Learning Algorithms to process between queries and to learn from data.
- It uses hardcoded algorithm and database format that searches from database file and answer the response corresponding to the query [if present in the database].

II. Methods and Algorithms

All these Methods are self coded or are derived form external libraries.

The flow of program is shown in Fig.4

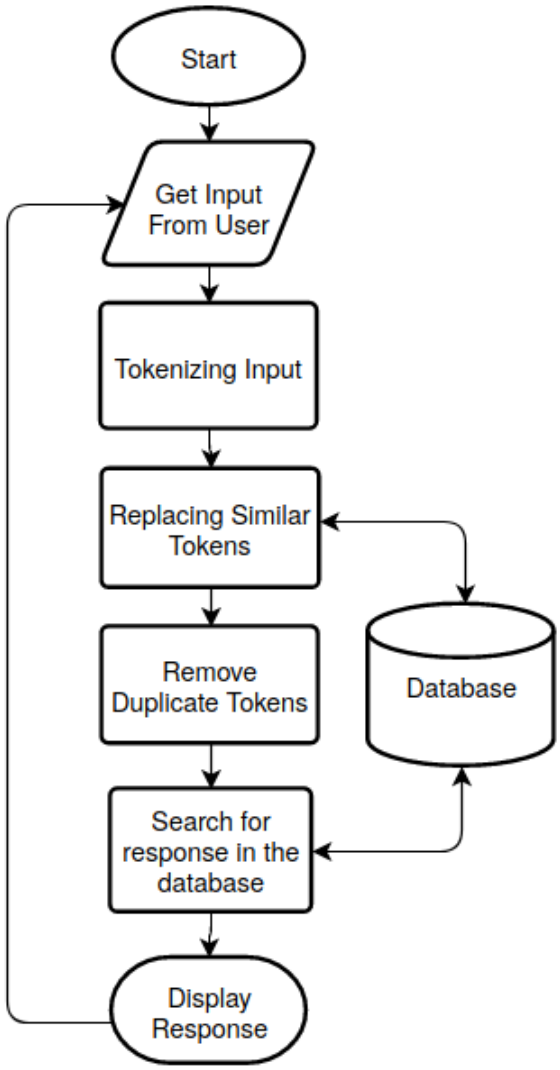

Fig. 4. Program Flow

A. Creating Class Objects

Chat-Bot Class object can be created using the bot or cpp-bot class. When creating new object `Initializing CHAT BOT` is shown on the screen.

B. Get Input

This app includes both GUI and TUI. Graphical User Interface uses HTML5, CSS and BOOTSTRAP for Web GUI; Flask for http server and Socket.IO for websocket connection to the backend app server.

To Explain the backend program, Python standerd Text User Interface is used, and will only be explained under this document. The chat bot waits for the user to input a query by using standerd Python `input()` function.

Python Standerd Text User Interface is shown in Fig. 5

```
YOU: what are loops in functions???

CPP_Bot: Loops can execute a block of
code as long as a specified condition
is reached.
```

Fig. 5. Text User Interface

C. Tokenizing with NLTK

NLTK (Natural Language Tool Kit) is widely used NLP (Natural Language Processing) Library which uses machine learning techniques for Natural Language Processing. This library is used for advance text mining and chatbots. In this program, we have used `word_tokenize()` function from this library that uses `punkt` for NLTK tokenizer training and seperate the words return a list(array) of words and symbols, which the AI thinks a seperate word or symbol. As it uses machine learning techniques, it has large performance time than builtin standerd function shown in Fig. 6

```
Tokenizing Input with NLTK
7.488037109375 miliseconds taken
['what', 'are', 'loops', 'in', 'functions
', '?', '?', '?']
```

Fig. 6. NLTK Tokenizer

D. Tokenizing with Standerd Library

Python has a builtin `split(delimiter)` function that uses delimiter to split the query and returns a list (array) which can be used in further processes. It is easier and has smaller runtime as it is a builtin function, which can be used after whitelisting english alphabets. It can improve performance time than using NLTK Library, as shown in Fig. 7

```
Tokenizing Input with split()
0.077880859375 miliseconds taken
['what', 'are', 'loops', 'in', 'functions']
```

Fig. 7. Builtin Split Function

E. Replacing Similar Tokens

We have used a synonyms List in the database file that include all the words that are similar. If the tokenized word is in the list, then it is replaced by the tag corresponding to the token, which can be processed in next steps to search the query with the tag.loops and functions is replaced by loop and function respectively as shown in Fig. 8

```
Replacing Similar Tokens with synonym's tag
7.592041015625 miliseconds taken
['what', 'are', 'loop', 'in', 'function', '?', '?', '?']
```

Fig. 8. Replaced Similar Tokens

F. Remove duplicate Tokens

We have used a synonyms List in the database file that include all the words that are similar. If the tokenized word is in the list, then it is replaced by the tag corresponding to the token, which can be processed in next steps to search the query with the tag.

G. Searching for response

Response Searching from the database is the vital algorithm and the heart of this program. The tokenized input array is fed the function. It checks and matches the words of the array sequencially. The queries are searched from the database in sequential order which ever matches first the response is returned. The order of database is ranked that which needs to be first responseded. e.g. what are loops in functions? Query for function is ranked below as loops can be inside function. The response is taken randomly from the query response array by `random.choice()` function.

H. Printing of response

After the response is returned, builtin `print()` function displays the response on the screen.

III. Results

Some of the queries asked from the chat-bot is shown below.

`YOU` : aoa
`BOT` : Wa alaikum Aslam

`YOU` : what can you do
`BOT` : I answers queries about C++

`YOU` : what are loops in python
`BOT` : I can only tell you about c++

`YOU` : define loops and function??
`BOT` : I am not programmed to process two or more queries at a time..

`YOU` : what is operator overloading in classes in c++?
`BOT` : Operator overloading is the ability to give the operators a special meaning for the data type.

These queries and responses are written in the database. If all of the queries doesn't match the response is as follows:

`YOU` : what is google
`BOT` : Good Question but I can't tell you anything about it.

IV. Discussion and Summary

AI Based apps are becoming popular in modern years. They are being used in many different applications such as Chat-Bot, ecommerce, bussiness, personalization and self driving cars etc. Machine Learning is most recent form of Aritficial Intelligence. Chat Bots and Virtual Assistants are found in our daily life. Such as Google Assistant for android mobile phones, Cortana for Microsoft Windows, Siri by Apple and Alexa by Amazon. Logos of some popular Assistants are shown in the figure Fig. 9

Fig. 9. Popular Virtual Assistants

Assistants such as Google Assistant, Alexa and SIRI etc can integrated with IFTTT and many other services and work together to make our lives easier. Assistants can be used to do tasks such as ”Turn on the Lights”, ”Query related to any topic”, ”Turn on Thermostat” etc.

From the above discussion, it is concluded that chat-bots can be used in different applications such as search enigne, performing different tasks, helping us in our daily work and improving workflow of the user.

References

[1] https://www.w3schools.com/cpp/default.asp
[2] https://www.w3schools.com/python/default.asp
[3] https://medium.com/@UdacityINDIA/why-learn-python-4149e541d884