Bachelor of Technology in Computer Engineering

A Project Report on: ***Gini – The chatbot***

Prepared by:

Admission No.	Student Name
U14CO001	Rajul Nahar
U14CO050	Nupur Modi
U14CO063	Riddhi Vora
U14CO105	Asmita

Class : B.TECH. IV (Computer Engineering) 8th Semester

Year : 2017-2018

Guided by : Prof. Dhatri Pandya

Prof. Vandana Joshi

DEPARTMENT OF COMPUTER ENGINEERING

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY,

SURAT - 395 007 (GUJARAT, INDIA)

INDEX

CHAPTER I
INTRODUCTION

1.1 Objective:

We have tried to create a chat bot that interacts with user and gives information about the following:

- SVNIT college premises
- Tells a joke
- Thought of the day
- Summary of daily news category wise
- Appropriate responses to spontaneous questions

1.2 Motivation:

In 1950, Alan Turing's famous article "Computing Machinery and Intelligence" was published, which proposed what is now called the Turing test as a criterion of intelligence. This criterion depends on the ability of a computer program to impersonate a human in a real-time written conversation with a human judge, sufficiently well that the judge is unable to distinguish reliably—on the basis of the conversational content alone—between the program and a real human. The notoriety of Turing's proposed test stimulated great interest in Joseph Weizenbaum's program ELIZA, published in 1966, which seemed to be able to fool users into believing that they were conversing with a real human.[1]

The classic historic early chatbots are ELIZA (1966) and PARRY (1972). While ELIZA and PARRY were used exclusively to simulate typed conversation, many chatbots now include functional features such as games and web searching abilities.

Chatbots are commonly found in messaging apps such as Facebook, telegram and in mobile phones as personal voice assistant such as Siri, Cortana and google assistant. Many banks and insurers, media and e-commerce companies, airlines and hotel chains, retailers, health care providers, government entities and restaurant chains have launched their own chatbots to answer simple questions, increase end customer engagement, promote their products and services, and give their customers a more convenient and easier way to order from them.

Our chatbot tries to give information about SVNIT to interested people, hence automating task of question answering.

CHAPTER II

IMPLEMENTATION AND TECHNOLOGY

2.1 Module description:

A chatbot can be considered as a question-answer system where experts provide knowledge for solicitation of user. A chatbot is software designed to simulate an intelligent conversation with a human partner.

The chatbot attempts to give answer to user's question regarding the college, Sardar Vallabhbhai National Institute of Technology. It is used to automate the task of answering queries to any visitor at the college premises. Additionally, it can also make light hearted jokes, provide thought of the day and tell latest news.

The chatbot is fundamentally a pattern matching program. It makes use of regular expressions. A regular expression (regex or regexp for short) is a special text string for describing a search pattern. It checks the user input against a list of regular expressions provided and upon finding the match, gives the output corresponding to that regular expression.

It contains a dictionary, reflections which maps first-person pronouns to second-person pronouns and vice-versa. It is used to "reflect" a statement back against the user. In this project, the user can ask the chatbot questions regarding the college like information about the departments, offices, hostels, dispensary and various other buildings. The chatbot can also tell jokes, thought of the day and provide latest news from BBC.

To tell the latest news, the module uses online news feeds and extracts news using a HTML parser (BeautifulSoup). After extracting the text, it eliminates the stopwords and calculates the frequency of all the words in the text [6]. The sentences containing words with frequency greater than threshold frequency are displayed as part of summarised text.

2.2 Technology used:

The programming language used is python. To incorporate the use of various natural language processing features like regular expressions, tokenization, stopword removal and pattern matching, the Natural Language Processing Toolkit(NLTK) is used [7].

2.2.1. NLTK

NLTK is a leading platform for building Python programs to work with human language data [2]. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum [3].

It was developed by Steven Bird and Edward Loper in the Department of Computer and Information Science at the University of Pennsylvania.

NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

2.2.2. Beautiful Soup

To extract latest news article from the web pages, Beautiful Soup is used. It is a Python library for pulling data out of HTML and XML files. It works with your favourite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree.

2.2.3. Collections: This module implements specialized container datatypes providing alternatives to Python's general purpose built-in containers, **dict**, **list**, **set**, and **tuple**[4].

Class collections.defaultdict([default_factory[,..]])

Returns a new dictionary-like object. **defaultdict** is a subclass of the built-in **dict** class. It overrides one method and adds one writable instance variable. The remaining functionality is the same as for the **dict** class and is not documented here.

The first argument provides the initial value for the **default_factory** attribute; it defaults to None. All remaining arguments are treated the same as if they were passed to the **dict** constructor, including keyword arguments.

2.2.4. urllib2

The **urllib2** module defines functions and classes which help in opening URLs (mostly HTTP) in a complex world — basic and digest authentication, redirections, cookies and more.

The **urllib2** module defines the following functions: [5]

urllib2.urlopen(*url [, data [, timeout [, cafile [, capath [, cadefault [, context]]]]]*)
Open the URL *url*, which can be either a string or a Request object.

data may be a string specifying additional data to send to the server, or None if no such data is needed. Currently HTTP requests are the only ones that use *data*; the HTTP request will be a POST instead of a GET when the *data* parameter is provided. *data* should be a buffer in the standard *application/x-www-form-urlencoded* format. The urllib.urlencode() function takes a mapping or sequence of 2-tuples and returns a string in this format. urllib2 module sends HTTP/1.1 requests with Connection:close header included.

The optional *timeout* parameter specifies a timeout in seconds for blocking operations like the connection attempt (if not specified, the global default timeout setting will be used). This actually only works for HTTP, HTTPS and FTP connections.

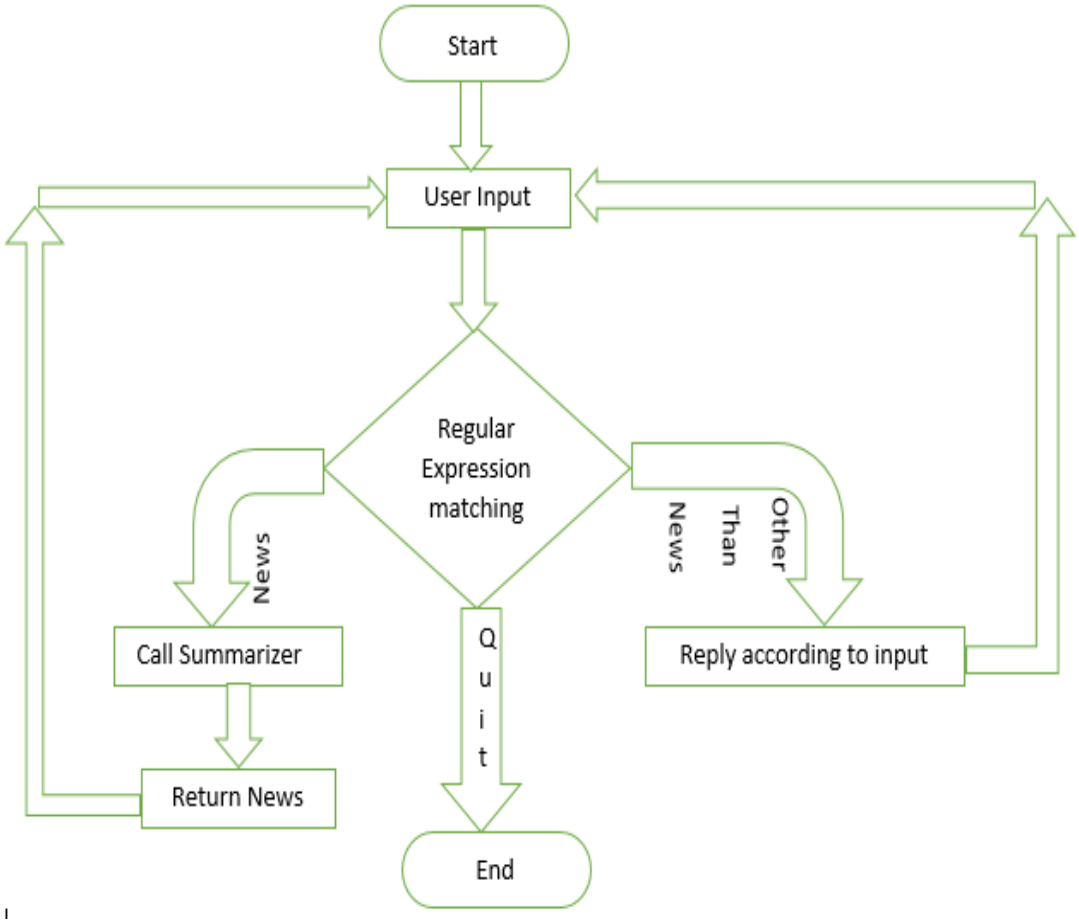
If *context* is specified, it must be a ssl.SSLContext instance describing the various SSL options. See HTTPSConnection for more details.

The optional *cafile* and *capath* parameters specify a set of trusted CA certificates for HTTPS requests. *cafile* should point to a single file containing a bundle of CA certificates, whereas *capath* should point to a directory of hashed certificate files. More information can be found inssl.SSLContext.load_verify_locations().The *cadefault* parameter is ignored.

This function returns a file-like object with three additional methods:

- geturl() — return the URL of the resource retrieved, commonly used to determine if a redirect was followed
- info() — return the meta-information of the page, such as headers, in the form of an mimetools.Message instance (see Quick Reference to HTTP Headers)
- getcode() — return the HTTP status code of the response.

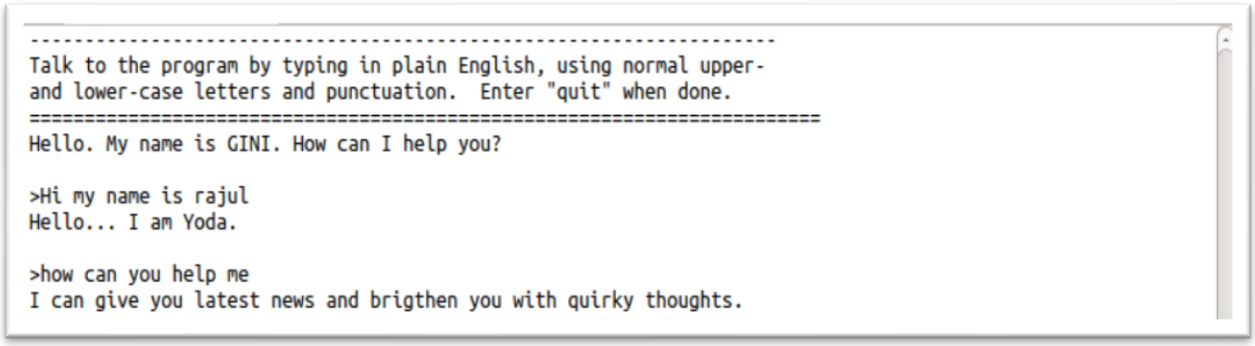
2.3 Flow Diagram:



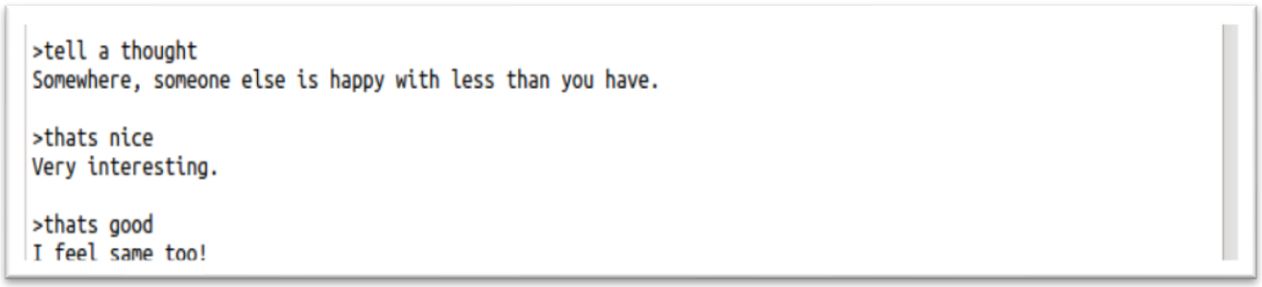
CHAPTER III

SIMULATION AND RESULTS

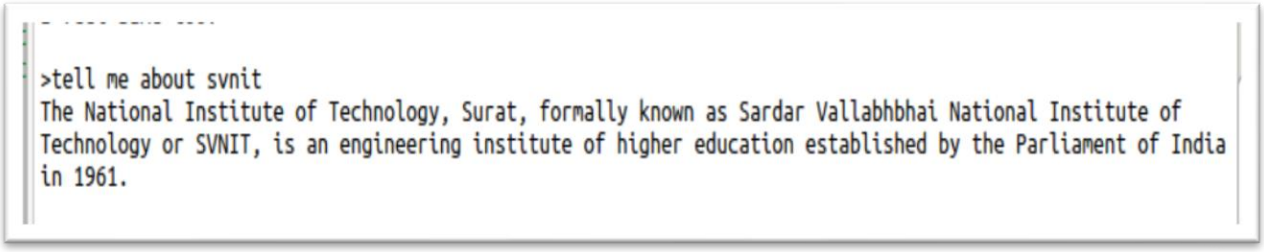
Snap of simulation with description



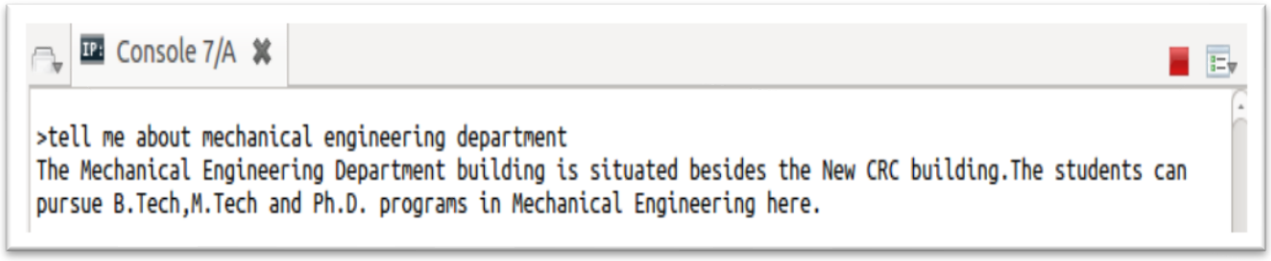
This is how the chatbot starts. Then the user has entered “Hi my name is rajul” to which the bot replied. The bot replies to the user telling how it can help.



The bot replies a thought when asked by the user.The bot is user friendly to all the responses.

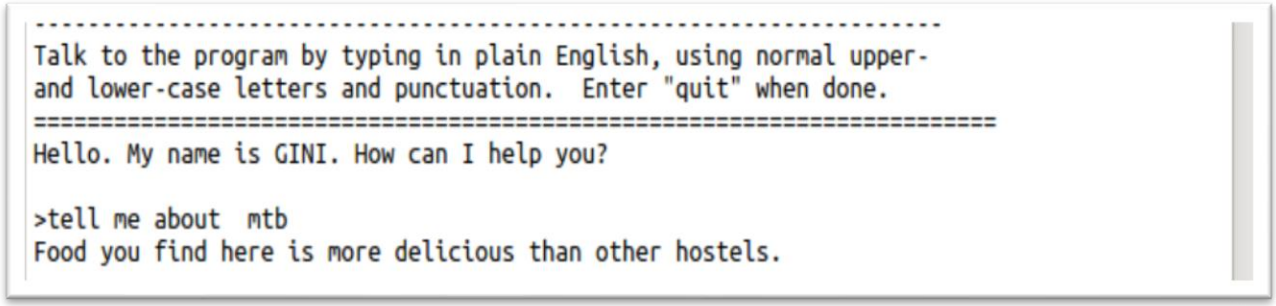


The bot helps the user with information regarding SVNIT.

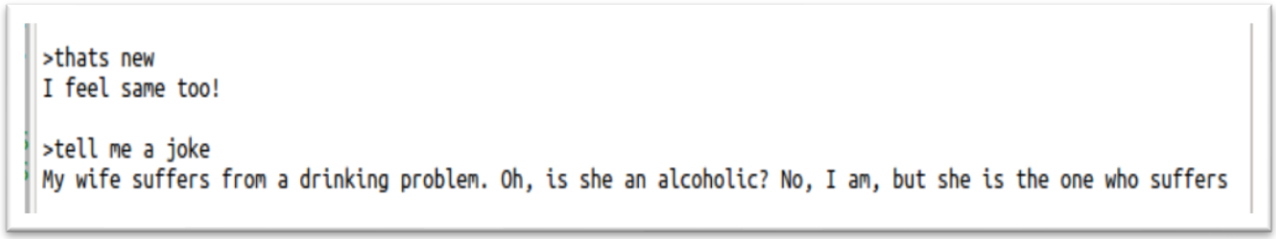


The bot gives information regarding departments when asked for a particular department.

It also provides information regarding different standalone buildings in our college like tnp, canteen, dispensary etc.



The bot provides information about hostels when asked accordingly.



The bot entertains the user with a joke.

```
>why should I JOIN SVNIT?

>Because I am made by students studying over there :P
TEHJHJHJ
Let's change focus a bit... Ask me something else.
```

The bot gives reply to a question asked specifically about SVNIT.

```
>tell me news
You can also ask like news about science and environment etc

>news about health
Parents facing 'unfair child abuse claims' over bruising - BBC News
 Share this with Email Facebook Messenger Messenger Twitter Pinterest WhatsApp LinkedIn Copy this link
These are external links and will open in a new window Parents are being investigated for possible child
abuse because of the misinterpretation of guidelines on bruising in babies, it's claimed.

>news about arts
Very interesting.

>news about technology
'Facebook in PR crisis mode', says academic at heart of row - BBC News
During the committee hearing, he explained that he was approached by SCL - the parent firm of Cambridge
Analytica - in spring 2014 about monetising an app he had developed at the University of Cambridge's
Psychometrics Centre.
```

The bot notifies the user regarding various categories of news it can provide.

The bot provides news to the user of particular category when asked for it.

CHAPTER IV

CONCLUSION AND FUTUREWORK

4.1 Drawbacks

The chatbot has some limitations: -

- Limited number of news categories. It cannot provide news about specific topic sbut instead gives news according to broad categories like sports, science, health and business.
- The chatbot contains limited information about college and hence cannot answer all types of queries.
- Limited dictionary of thought and jokes.

4.2 Conclusion

We have satisfactorily implemented the chat bot for our purpose. It provides information regarding college, entertains the user and also gives news about some specific categories.It is quite user friendly. It succesfully automates the task of answering questions of people seeking information about college.

Although it is limited by the type and number of queries asked, it still provides answers to the most frequent questions asked and hence can serve as satisfactory query answering system for college.

4.3 Future Work

The implementation of chatbot is quite basic. It does not learn from its experiences. The type of queries are limited due to limited use of regular expressions.

Machine Learning can be used to enhance features and more functionalities can be added. It can have various advanced features that are observed in Siri by Apple or Cortana by Microsoft.

REFERENCES

[1] https://en.wikipedia.org/wiki/Chatbot

[2] https://www.nltk.org/

[3] http://www.nltk.org/api/nltk.chat.html

[4] https://docs.python.org/2/library/collections.html

[5] https://docs.python.org/2/library/urllib2.html

[6] https://www.crummy.com/software/BeautifulSoup/bs4/doc/

[7] https://medium.com/@allanmeriales/a-simple-chatbot-using-nltk-chat-640456dcdf72