IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

AI Chatbot

Mr. Mayank Raj[1], Rishabh Kumar Jha[2], Ashmit Tiwari[3], Jaideep Singh[4], Uday Singh[5]
Assistant Professor, Department of Computer Science & Engineering[1]
Students, Department of Computer Science & Engineering[2,3,4,5]
I.T.S Engineering College, Greater Noida, India

Abstract: *There has been an emerging trend of an enormous number of chat applications which are present within the recent years to helppeople to connect across different mediums, like Hike, Whats App, Telegram, etc. The proposed network-based android chat application used for chatting purposes with remote clients or users connected to the online , and it will not let the user send inappropriate messages. This paper proposes the mechanism of creating knowledgeable chat application which can not permit the user to send inappropriate or improper messages to the participants by incorporating the bottom level implementation of natural language processing (NLP). Before sending the messages to the user, the typed message evaluated to hunt out any inappropriate terms within the message which may include vulgar words, etc., using natural language processing. The user can build their own dictionary which contains vulgar or irrelevant terms. After pre-processing steps of removal of punctuations, numbers, conversion of text to lower case and NLP concepts of removing stop words, stemming, tokenization, named entity recognition and parts of speech tagging, it gives keywords from the user typed message. These derived keywords compared with the terms within the dictionary and database and therefore the respective response are replied.*

Keywords: Android Application, Chatting, Dictionary, Named Entity Recognition, Natural Language Processing, Networking, Parts of Speech tagging, Sentimental Analysis, Stemming and Tokenization

I. INTRODUCTION

A chatbot could even be a program that simulates interactive human conversation by using key pre-calculated user phrases and auditory or text-based signals. Chatbots are frequently used for basic customer service and marketing systems that frequent social networking hubs and instant messaging clients. they're also often included in operating systems as intelligent virtual assistants.

Modern chatbots are frequently utilized in situations during which simple interactions with only a limited range of responses are needed. this will include customer service and marketing applications, where the chatbots can provide answers to questions on topics like products, services or company policies. If a customer's questions exceed the talents of the chatbot, that customer is usually escalated to an individual's operator.

Online chatting refers to the tactic of sending and receiving messages using the online . There are various chatting applications available within the market. At the primary quarter of 2017, the entire number of users using chat applications are quite 5.03 Billion [1], and widely used apps are WhatsApp, Facebook Messenger, We Chat, QQ Mobile, etc., of those applications provide various features to form sure security, integrity, and consistency. of these apps let the user send any messages, and thus the messages are often lewdor inappropriate. There are many cases filed for sending lewd or inappropriate messages in various online mediums [2,3,4,5,6]. it's going toeven be possible for the user to send inappropriate messages by mistake.

Section 66A of the knowledge Technology (Amendment) Act, 2008 says that transmitting obscene information using transmission equipment which may end in three years of incarceration including a fine [7,8]. to unravel these concerns, the proposed mechanism implemented.

The proposed network-based android chat application used for chatting purposes with remote clients or users connectedto the online , and it will not let the user send inappropriate messages. the appliance developed for Android, because Android is one of the foremost widely used mobile operating systems having major market share as compared to other mobile operating systems like iOS, Windowsand Blackberry [9,10].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

The main objective of the project is to build a chatbot using a database that would understand and reply to student queries byimporting Natural Language Processing Tool Kit (NLTK). The student should get replies to his queries regarding his details like his attendance or his Academic percentages and also the details about the college like cultural Fests, Flash Mobs, Technical Fests, and Workshopsrelated Details

By creating this chatbot if would be helpful for the student and also to the respected Parents to easily find out details or any information regarding the college. The student would not have to go to the management for minor information which he can get by the chatbot.

II. LITERATURE SURVEY

ELIZA: Eliza is taken under consideration to be the first chatbot within the history of computing which was developed by Joseph Weizenbaum at Massachusetts Institute of Technology (MIT). It was in 1994 that the term “Chatterbot" was coined. ELIZA operates by recognizing keywords or phrases from the input to reproduce a response using those keywords from pre-programmed responses. For instance, if a human says that “My mother cooks good food”. ELIZA would devour the word “mother”, and respond by asking an open- ended question “Tell me more about your family”. This created an illusion of understanding and having an interaction with a true person through the method was a mechanized one.

Another chatbot named Alice. It was developed in 1995 by Richard Wallace. Unlike Eliza, Alice chatbot was ready to usenatural language processing, which allowed for more sophisticated conversation. It was revolutionary, though, for being open-source. Developers could use AIML (artificial intelligence markup language) to create their chatbots powered by Alice.

ALICE: The Bot That Launched a Thousand other Bots: No list of innovative Chatbots would be complete without mentioning ALICE, one of the very first bots to travel online and one that's delayed incredibly well despite being developed and launched quite 20 years ago. ALICE

-which stands for Artificial Linguistic Internet Computer Entity, an acronym that would are lifted straight out of an episode of The X-Files

,was developed and launched by creator Dr. Richard Wallace way back within the dark days of the first Internet in 1995. (As you'll see within the image above, the website's aesthetic remains virtually unchanged since that point , a strong reminder of how far web design hascome.)

Even though ALICE relies on such an old codebase, the bot offers users a remarkably accurate conversationalexperience. Of course,no bot is ideal , especially one that's sufficiently old to legally drink the U.S. if only it had a physical form. ALICE, like many contemporarybots, struggles with the nuances of some questions and returns a mix of inadvertently postmodern answers and statements that suggest ALICE has greater self-awareness for which we might give the agent credit. For all its drawbacks, none of today's chatbots would are possible without the groundbreaking work of Dr. Wallace. Also, Wallace's bot served because the inspiration for the companion OS in Spike Jonze's 2013 science-fiction romance movie, Her.

Jabberwacky may be a chatterbot created by British programmer Rollo Carpenter. Its main aim is to "simulate natural human chat during a stimulating , entertaining and humorous manner". It is an early attempt at creating AI through human interaction. The stated purpose of the project was to make AI that's capable of passing the Turing Test. It is designed to mimic human interaction and to hold out conversations with users. It is not designed to hold out the other functions. Unlike more traditional AI programs, the training technology is meant as a sort of entertainment instead of getting used for computer support systems or corporate representation.

Recent developments do allow a more scripted, controlled approach to take a seat atop the overall conversational AI, getting to compile the simplest of both approaches, and usage within the fields of sales and marketing is underway. Its creator believes that it are oftenincorporated into objects around the home like robots or talking pets, intending both to be useful and entertaining, keeping people company.

Mitsuku may be a chatbot created from AIML technology by Steve Worswick. It claims to be an 18-year-old female chatbot from Leeds, England. It contains all of Alice AIML files, with many additions from user-generated conversations, and is usually a piece ongoing .

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

For example, if someone asks "Can you eat a house?" , Mitsuku looks up the properties for "house". Finds the worth of "made from" is about to "brick"and replies "no", as a home is not edible. She can play games and do magic tricks at the users request. In 2015 sheconversed, on average, more than a quarter of a million times daily.

Current chatbots are developed using a variety of methods like rule-based where rules are hard-coded in code, AI-based bots, pattern-basedwhich can handle only mentioned patterns for retrieving answers. There are frameworks available for developing chatbots but they also use either rule-based or pattern-based techniques. In rule-based chatbots that are easiest to build, one needs to write rules like If X then Y else if A then B, etc. So if there are 100 scenarios, the developer needs to write 100 rules for each of the scenarios. The volume, variety, and complexity of data make such techniques inefficient.

It is nearly impossible to write rules and/or patterns for massively available data. AI-based bots are built on NLP and ML. They are based on the human capability of learning information but with more efficiency. Natural Language Processing (NLP) are often used where predefined or static rules, patterns might not work.

On comparing to all of these chatbots a special chatbot system is to be proposed for each enterprise like banking-related chatbots, educational related chatbots, medical-related chatbots, etc.., which works on that enterprise more effectively with good efficiency within a short time. For this kind of chatbots, we have to incorporate the Natural Language Processing (NLP) which is used to resolve language-related error queries and acts according to it and removes the sentimental barriers and makes the conversation effectively in a good way.

III. PROPOSED SYSTEM

During this proposed system, we focus mainly on the failures to spot the inappropriate context within the text message and is that the main reason for various problems. the prevailing work states that there are not any methods or provisions for the identification of lewd or vulgar context within the typed messages. The Proposed system solves these issues by developing an NLP based android tool that identifies and warns the user if the sentiment of the message contains a lewd or vulgar context.

the essential approach for developing this chatbot is to find out and importing NLTK which contains the text processing libraries, MySQL which is an open-source database, SKlearn 0.0 which may be a library in python that gives many unsupervised and supervised learning algorithms. it's on NumPy, SciPy, and matplotlib, this library contains tons of efficient tools for machine learning and statistical modeling including classification, regression, clustering and conditionality reduction.

During this Proposed method, we are getting to provide separate login id and password as login credentials to oldsters and students which is merely operable by the administrator. Upon successful login, a chatbot conversation is initiated on the interface which accepts the queries and responds in consistent with it. If the questions exceed the skills of the chatbot, that customer is automatically escalatedto a person.

IV. SYSTEM ARCHITECTURE

The primary phase deals with data pre-processing. The results of the primary step given as input to the second phase. The second phase deals with the implementation of NLP concepts just like the removal of stop words, stemming, entity recognition, tokenization and parts of speech tagging which derive keywords from the user typed the message. These keywords compared with the user dictionary to spot irrelevant terms. The third phase deals with sending and receiving the messages using the web and saving the messages in encrypted form within the real-timedatabase "MySQL".

The NLP based tool implemented using Python, and MySQL. The tool accepts the user typed message as an input. Data pre- processing applied to the user typed message. Data pre-processing may be a procedure of knowledge mining, executed on real-time data, which is vociferous, inadequate or erratic. the next data scrubbing measures are employed :

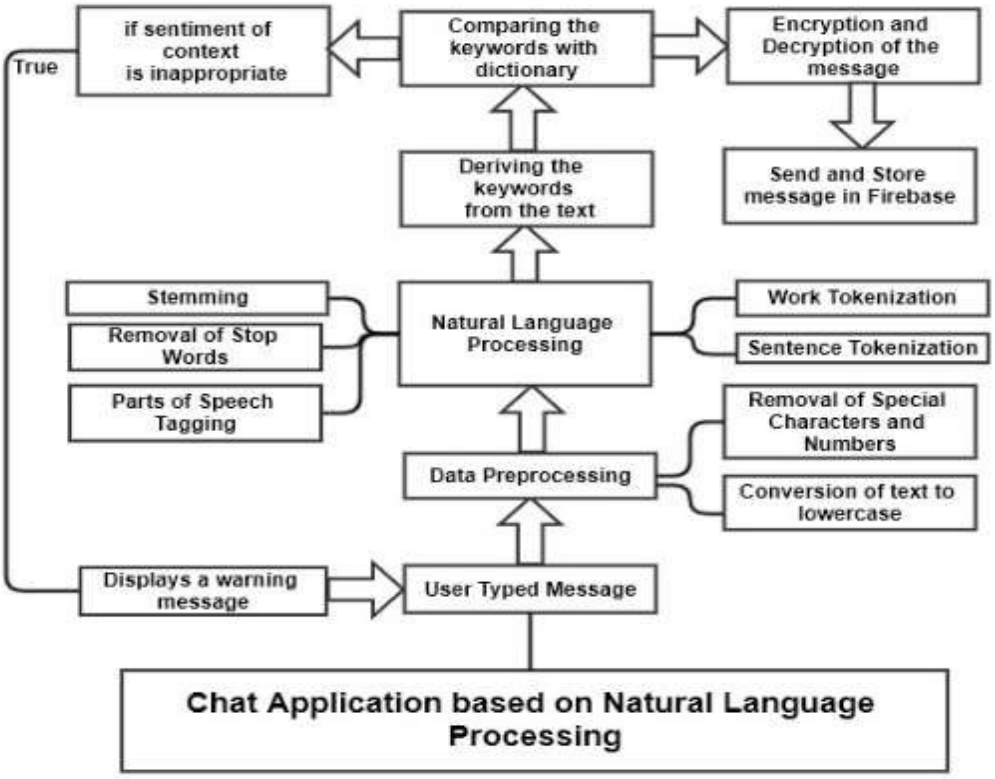
Eliminating all the distinctive letters from the line .

Trimming undesired tabs, spaces, newlines and extra nonprintable letters within the line.

Transforming the whole text to small letter .

All these steps performed to form computation easy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023



Natural Language Processing

V. IMPLEMENTATION

NLTK stands for natural language Toolkit. This toolkit is one of the foremost powerful NLP libraries which contains packages to make machines understand human language and reply thereto with an appropriate response. Tasks performed by NLTK are: Automatic text summarization, Translation, Named entity recognition, relationship extraction, and sentiment analysis.

Natural language processing concepts applied to the pre processed data. natural language Processing could also be a way for Machines to guage , determine, and acquire the semantics of human language wisely. Python programming language is used to accomplish several tasks of natural language Processing on cleaned data. Word Tokenization, Sentence Tokenization, Removal of Stop words, Stemming, Entity Recognition, Parts of Speech Tagging, etc. are practiced to reconstruct the data to a pattern suitable for interpretation.

Word Tokenization is that the tactic of adjusting the text into tokens and saving it within the list. Sentence tokenization is that the tactic of transforming the text into different phrases. Stop words deemed inapplicable or pointless because they need limited importance in capturing the semantics of the text and additionally stop words increase the searching time which finishes up in wastage of the many computational resources. Stop words are omitted to preserve both time complexity and space complexity. Stemming could also be a crude method that cuts off the ends or beginning of words. the first aspiration of stemming is to vary a derivative word into its standard form and keeps the idea word. Parts of Speech Tagging examines the text and assigns parts of speech to each token as a verb, adjective, noun, etc., Entity Identification helps to classify the named entities like persons, organizations, etc., from the text. After implementing NLP techniques on the pre-processed cleaned data, definitive keywords derived. These keywords compared with the keywords available within the user dictionary. The vocabulary contains all the keywords. From the keywords procured from the text, the sentiment of the context determined. If the meaning of the message is in appropriate, then user not allowed to send the message.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

MODULES

Tokenization

Tokenization is that the process by which big quantity of text is split into smaller parts called tokens. Natural language processing is employed for building applications like Text classification, intelligent chatbot, sentimental analysis, language translation, etc. It becomes vital to knowthe pattern within the text to realize the above-stated purpose. These tokens are very useful for locating such patterns also as is taken into account as a base step for stemming and lemmatization. For the notice , don't be concerned about stemming and lemmatization but treat themas steps for textual data cleaning using NLP (Natural language processing).

Stemming

Stemming is that the process of manufacturing morphological variants of a root/base word. Stemming programs are commonly mentionedas stemming algorithms or stemmers. A stemmer reduces the words “chocolates”, “ chocolatey”, “choco” to the basis word, “chocolate” and “retrieval”, “retrieved”, “retrieves” reduce to the stem “retrieve.

Lemmatization

Lemmatization is that the process of grouping together the various inflected sorts of a word in order that they are often analyzed as one item.Lemmatization is analogous to stemming but it brings context to the words. So it links words with similar aiming to one word.

Text pre processing includes both Stemming also as Lemmatization. repeatedly people find these two terms confusing. Some treat these two as same. Lemmatization is preferred over Stemming because lemmatization does the morphological analysis of the words.

Punctuation

Removing punctuations from the raw text may be a processing technique, which is employed to wash and prepare the text for modellingwith machine learning. After pack up the text is completed then each character and words are checked with the stored words within the database. so, if they're matching then the chatbot accepts it as a meaningful query and reply consistent with it.

Parts of speech(POS) Tagging

Tagging the method of classifying words into their parts of speech and labelling them accordingly is understood as part-of-speech tagging orjust tagging. Parts of speech also are referred to as word classes or lexical categories. the gathering of tags used for a specific task is understoodas a tag set.

VI. RESULTS AND DISCUSSION

Testing is to be done to discover errors, upon testing we get the results on giving the input such as text. To understand clearly, followthe below examples:

1) On entering the wrong password or wrong id.

```
Enter user id:16B81A1201
enter password:qwerty
u have entered wrong password
>>>
```

2) Upon successful login, a chatting session is initiated.

```
Enter user id:16B81A1203
enter password:16B81A1203
 hai  ('Nreddy',) CHATBOT:  This is cvr college CHATBOT. I will answer your queries about College details. If you want to exit, type Bye!
Enter query:hi
CHATBOT: hi
```

Once the login is done, chatbot initiates a chat session, the user can ask any number of queries to collect the information from the chatbotsuch as academic percentages, attendance percentages, etc..,on entering any irrelevant or unmeaningful query as input, it generates the below message.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

```
Enter query:utdfhq0wdi-29dk
CHATBOT: I am sorry! I don't understand you
Enter query:
```

VII. CONCLUSION

This project entitled "Human-Chatbot Interaction using NLTK." is beneficial for any enterprise to automatically initiating a talk with thepurchasers and receiving the queries and responding consistent with those queries. mainly this type of chatbot is beneficial to get rid of thebarrier of employing accurate grammar because we are using a natural language processing which solves the grammar-related issues and solves that queries, and also uses stop words. It detects the bad words and counting on an "inappropriate words" warning message, we are filtering out useless data. This project eradicates the sentimental words too. This Project helps finally for the development of a chatbot for an enterprise effectively and with good efficiency

REFERENCES

[1]. Aafiya Sheikh, Dipti More, Ruchika Puttoo, Sayli Shrivastav, Swati Shinde-"A Survey paper on chatbots" 2019, IRJET-V614383.

[2]. Karthik S, R John Victor, Manikandan S, Bhargavi Goswami-"Proffesional Chat Application based on Natural Language Processing" 2017, IEEE.

[3]. Bhaumik kohli, Tanupriya Choudhary, Shilpi Sharma, Praveen Kumar-"A Platform Human-Chatbot interaction using python" 2018, IEEE Conference.

[4]. Jennifer Hill, W.Randolph Ford,Ingrid G..Farreras-"Real conversations with artificial intelligence: A comparison between human- human online conversations and human-chatbot conversations" 2015, ELSEVIER.

[5]. https://scikit-learn.org

[6]. http://www.nltk.org