RESEARCH ARTICLE OPEN ACCESS

CHATBOT FOR MOVIE APPLICATIONS

Nalluri Reshma[1], Nerella Lakshmi Tejaswi[2], Pothineni Gayathri Sowjanya[3], Peddu Charitha[4], K Vikas[5]

[1, 2, 3,4]B-Tech, Dept of CSE, VVIT, Guntur, AP.
[5]Assistant Professor, Dept of CSE, VVIT, Guntur, Ap.

***Abstract*——** Chat Bots are the programs that attempts to simulate the conversation of human being via text or voice interactions using Artificial
Intelligence Markup language(AIML). Artificial Intelligence Markup Language (AIML) is a way of making a computer-controlled robot, or a software think intelligently, in the similar manner the humans think.In this paper we provide the design of a ChatBot for movie application where user can get complete details for what they are looking for , which provides an efficient and accurate answer for any query based on the dataset of FAQs using AIML. Template based and general questions like welcome/greetings and general questions will be responded using AIML that will serve user need. Latent Semantic Analysis(LSA) is one of the component of Natural Language Processing(NLP).LSA is the ability of a computer program to understand human language
. LSA allows ChatBots to understand the messages and respond appropriately. This ChatBot is prepared to get movie details and answer FAQs in an interactive fashion. .

Keywords—Artificial Intelligence Markup Language(AIML), Latent Semantic Analysis(LSA), Pattern Matching, Chatbot, Human Computer Interaction(HCI)

•

I. INTRODUCTION

Now-a-days people are attracting too much for entertainment by watching movies and due to time constraints and their busy schedule they are willing to
complete all the tasks with in a short span of time , but searching for different movies in different websites is time consuming and is of long process .To overcome the above problem we creating an efficient Bot which interact with the user in friendly manner. This bot for movie application is created with the help of AIML and LSA which is a part of Natural Language Processing(NLP). AIML is an XML based mark up language used to create Artificial intelligence applications. Latent semantic analysis is used for processing the query given by the user and compares with the existing questions present in the database and helps in retrieving the appropriate answer with minimal response time.

There are numerous websites available on World Wide Web which helps to host our personalized chatbot that can respond intelligently to human queries. These services are used by many industries, organizations or institutions to service their consumers. One of the most widely used language for the development of bot is AIML (Artificial Intelligence Markup Language). This proves to be a deterministic language in terms of development of chat bots. AIML or Artificial Intelligence Markup Lanuage is an xml base Language used for creating chatbot (AI),it is always saved with the extension(.aiml). The basic units of an AIML dialogs are called categories .A category consists of
(i) an user input ,in the form of a sentence
(ii) a response to user input ,presented by the chatterbot, (iii)an optional context.[3]

Almost all bots are developed using AIML in which all the

possible queries are enclosed in <pattern> and <template>
tags which contains question and answer respectively.

E.g.
<pattern>HOW ARE YOU</pattern> <template>Yeah! am fine. How are you?</template>

The patterns and template tags are further enclosed in
<category> tags. Thousands of categories tags are used to develop chatting bots. These categories are saved in database
when the bot is compiled or published. When a user sends a message the query pattern is matched with the pattern's stored in database and the corresponding template is sent as a response to the user.

II. Literature Review

This chapter gives a brief summary about all the researches
done until now. It also abstracts all research papers related to
Chatting Bots that have conducted till now with their technology used, enhancements in previous technology but the hurdles that the users are facing. Consequently, this formulates the problem statement of this project which explicates bot development and performance issues and proposed architecture to surmount it.

A. INTELLIGENT CHATBOT FOR EASY WEB-ANALYTICS INSIGHTS

In this fast-moving data-driven world, it is vital that we draw the accurate insights to make the right decisions at the right time. In terms of online websites, there are many web analytics tools that will give us performance reports. This paper compares two widely used analytic tools based on their ease of use.Once the bot-user types in the query in the chatbot, this chat bot will identify the category that contains the query pattern. Here the bot-user is expected to type in the query in a predefined pattern. Once the query pattern is matched, template of the category that contains the response is sent back to the bot-user. There are 3 query scenarios that can be considered and the scenarios are domain related,general queries and the one not belonging to these both.[6]

B. AN INTELLIGENT WEB-BASED VOICE CHAT BOT

This paper deals with the working of AIML based chat robot. A Java Program is developed which convert AIML files into database. This program is embedded into website which can in turns help its customers to develop bots. The major technological enhancement in this research is integrating speech recognition and text to speech converter. This empowers the bots to respond to user queries using voice instead of text and humans to chat with bots using voice instead of text messages.[8]

C. CHATBOT FOR UNIVERSITY RELATED FAQS

This paper develops an interactive chatbot for University related Frequently Asked Questions (FAQs), and the work flow of proposed framework. User discussion as a rule begins with welcome or general questions. User inquiries are first taken care by AIML check piece to check whether entered inquiry is AIML script or not. AIML is characterized with general inquiries and welcome which is replied by utilizing AIML formats.

This will help the student to fetch information like ranking of university, availability of services, university environment, updates regarding activities happening inside campus and many more and other academic information.[5]

D. DEVELOPMENT AND IMPLEMENTATION OF A CHATBOT IN A SOCIAL NETWORK

This paper describes the linking of chat bot with social network. It describes that how a chat bot can be linked with Twitter to entertain the users. It can also be used for advertisements. The bot is linked with Twitter since it part from a simple concept, the exchange of short messages no longer than 140 characters which drastically reduces the amount of information and the way it is published. The algorithm process in this bot is divided into three different parts:

Message reception. Message processing.
Generation of a suitable reply.[7]

III. PROPOSED METHOD

Failure of getting correct results of movies for which we are searching for in short span of time in different websites and Failure to identify the inappropriate context in the text
message is the main reason for introducing of this chatbot for movie application. The proposed system solves these issues by developing a bot using NLP and LSA which identifies and warns the user if the sentiment of the message contains lewd or vulgar context. [2]

Working of LSA consists of three phases.

- The first phase deals with data pre-processing. The result of the first step given as an input to the second phase.
- The second phase deals with implementation of NLP concepts 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 user dictionary to identify irrelevant terms.
- The third phase deals with sending and receiving the messages using the internet and saving the messages in encrypted form in the real-time database [4]

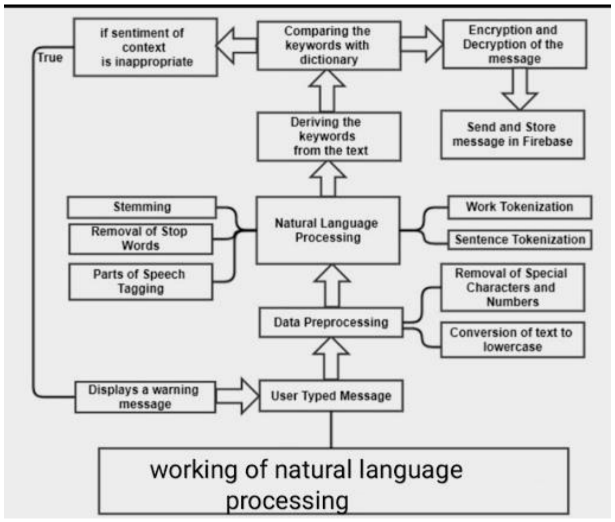
True
if sentiment of context is inappropriate
Comparing the keywords with dictionary
Encryption and Decryption of the message
Deriving the keywords from the text
Send and Store message in Firebase
Stemming
Removal of Stop Words
Parts of Speech Tagging
Natural Language Processing
Work Tokenization
Sentence Tokenization
Removal of Special Characters and Numbers
Data Preprocessing
Conversion of text to lowercase
Displays a warning message
User Typed Message
working of natural language processing

Figure:1 Working of NLP

Latent Semantic Analysis (LSA), also known as Latent Semantic Indexing (LSI) literally means analyzing documents to find the underlying meaning or concepts of those documents. If each word only meant one concept, and each concept was only described by one word, then LSA would be easy since there is a simple mapping from words to concepts.

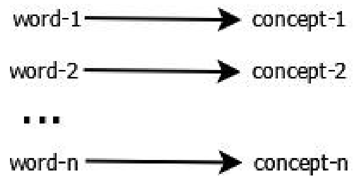

Figure:2

Unfortunately, this problem is difficult because English has different words that mean the same thing (synonyms), words with multiple meanings, and all sorts of ambiguities that obscure the concepts to the point where even people can have a hard time understanding.

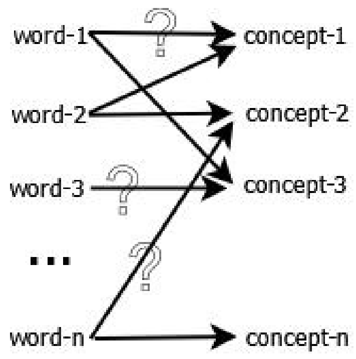

Figure:3

There are three types of AIML classes:
phrase result and would discover an association
1) Atomic categories amongst result and subject and it would react with
2) Default categories a question
3) Recursive Categories[1]

Atomic category : It is a sort of AIML classification The chatbot does not comprehend what it is stating. It just creates about this situation there was a control connecting family and mother. At that point it needed to create an answer that would urge the client to prevent another question with the trust of finding a keyword.
But more often than not ELIZA gives standard answers if a keyword where there in the exact match.

```
< category >
   < pattern > How are you
                                     < /pattern > <template> I am fine
      </template>
< /category >
```

In the above illustration if the client submits How are you then the chatbot will answer I am fine is not found .

Default category : It is a kind of AIML class where the utilization of a wild card symbols, for example, * is utilized to coordinate any info

```
        <category>
    < pattern >
            Who is * < /pattern >
             < template >
                 He is my brother
             < /template >
        < /category>
```

In the event that ALICE does not discover an answer utilizing the main classification it continues to the next.

Recursive category : It is a type of an AIML classification to persuade the user to be more particular.

```
        < category > <pattern>
Can you tell who the * is </pattern>
   < template > He is my brother
        < srai > Who is * < /srai > </template>
< /category >
```

Alice: Hi there! User: How old are you? Alice: I am eighteen in computer years
For such situation the wild symbol * is utilized to decrease User: Who is your father? Alice: I was programmed by Dr.Richard S. Wallace the information presented by user.
User: so you are a software program? Alice: Interesting deduction. Thanks and you are a
A. AIML Tags delightful person too. Am I a software program? User: wonderful! So what's your favorite food ALICE? [5]

IV. IMPLEMENTATION

The first step in Latent Semantic Analysis is to create the word by title (or document) matrix. In this matrix, each index word is a row and each title is a column. Each cell contains the number of times that word occurs in that title. For example, the word "book" appears one time in title T3 and one time in title T4, whereas "investing" appears one time in every title. In general, the matrices built during LSA tend to be very large, but also very sparse (most cells contain 0). That is because each title or document usually contains only a small number of all the possible words. This sparseness can be taken advantage of in both memory and time by more sophisticated LSA implementations.
In the following matrix, we have left out the 0's to reduce clutter.

STEPS FOR PROCESSING:

STEP 1 : Import Functions

STEP 2 : Define Data

Next, we define the data that we are using. Titles holds the 9 book titles that we have gathered, stop words holds the 8 common words that we are going to ignore when we count the words in each title, and ignore chars has all the punctuation characters that we will remove from words. We use Python's triple quoted strings, so there are actually only 4 punctuation symbols we are removing: comma (,), colon (:), apostrophe ('), and exclamation point (!).

STEP 3 : Define LSA Class

The LSA class has methods for initialization, parsing documents, building the matrix of word counts, and calculating. The first method is the __init__ method, which is called whenever an instance of the LSA class is created. It stores the stop words and ignore chars so they can be used later, and then initializes the word dictionary and the document count variables.

STEP 4 : Parse Documents

The parse method takes a document, splits it into words, removes the ignored characters and turns everything into lowercase so the words can be compared to the stop words. If the word is a stop word, it is ignored and we move on to the next word. If it is not a stop word, we put the word in the dictionary, and also append the current document number to keep track of which documents the word appears in.
The documents that each word appears in are kept in a list associated with that word in the dictionary.
After processing all words from the current document, we increase the document count in preparation for the next document to be parsed.

STEP 5 : Build the Count Matrix

Once all documents are parsed, all the words (dictionary keys) that are in more than 1 document are extracted and sorted, and a matrix is built with the number of rows equal to the number of words (keys), and the number of columns equal to the document count. Finally, for each word (key) and document pair the corresponding matrix cell is incremented. def build(self):

STEP 6 : Print the Count Matrix

The printing method is very simple, it just prints out the matrix that we have built so it can be checked.

STEP 7: Test the LSA Class

After defining the LSA class, it's time to try it out on our 9 book titles. First we create an instance of LSA, called mylsa, and pass it the stopwords and ignore chars that we defined. During creation, the __init__ method is called which stores the stopwords and ignore chars and initializes the word dictionary and document count.

Next, we call the parse method on each title. This method extracts the words in each title, strips out punctuation characters, converts each word to lower case, throws out stop words, and stores remaining words in a dictionary along with what title number they came from.

Finally we call the build() method to create the matrix of word by title counts. This extracts all the words we have seen so far, throws out words that occur in less than 2 titles, sorts them, builds a zero matrix of the right size, and then increments the proper cell whenever a word appears in a title.
Here is the raw output produced by print method. As you can see, it's the same as the matrix that we showed earlier.
Here is the raw output produced by print method. As you can see, it's the same as the matrix that we showed earlier.

. Conclusion and Future Scope

A chatbot is a rising trend and chatbot increases the effectiveness of business by providing a better experience with low cost. A simple chatbot is not a challenging task as compared to complex chatbots and developers should understand and consider the stability, scalability and flexibility issues along with high level of intention on human language. In short, Chatbot is ecosystem and moving quite fast and with the passage of time new features are added in the existing platform.For future work algorithm can be implemented more accurately and voice based search can form a new trend in searching techniques.

REFERENCES

[1] AM Rahman1 , Abdullah Al Mamun1 , Alma Islam2," Programming challenges of Chatbot: Current and Future Prospective", IEEE Region 10
Humanitarian Technology Conference (R10-HTC),2017

[2] Yoko Nishihara, Masaki Ikuta, Ryosuke Yamanishi, and Junichi Fukumoto," A Generation Method of Back-channel Response to Let a Chatting bot Be a Member of Discussions in a Text-based Chat", 6th IIAI International Congress on Advanced Applied Informatics,2017

[3] Saqib G., Faizan K, Ghatte N," Intelligent Chatting Service Using AIML", IEEE International Conference on Current Trends toward Converging Technologies

[4] Karthick S1 , R John Victor2 , Manikandan S3 , Bhargavi Goswami4," Professional Chat Application based on Natural Language Processing"

[5] Bhavika R. Ranoliya , Nidhi Raghuwanshi and
Sanjay Singh," Chatbot for University Related FAQs"Ramya Ravi," Intelligent Chatbot for Easy Web-Analytics Insights"

[6] Salto Martínez Rodrigo, Jacques García Fausto Abraham, Development and Implementation of a Chat Bot in a Social Network, in Ninth International Conference on Information Technology - New Generations, 2012, p. 751.

[7] Salomon Jakobus du Preez, Manoj Lall and Saurabh Sinha, An Intelligent Web-Based Voice Chat Bot, in EUROCON 2009, EUROCON '09. IEEE, 2009, p. 386.