International Journal of Computer Engineering & Technology (IJCET)
Volume 8, Issue 5, Sep-Oct 2017, pp. 15–22, Article ID: IJCET_08_05_003
Available online at
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=8&IType=5
Journal Impact Factor (2016): 9.3590(Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6367 and ISSN Online: 0976–6375
© **IAEME** Publication

WHATSAPP AUTO RESPONDER USING NATURAL LANGUAGE PROCESSING AND AI

Y. Nagender
Asst. Professor, CSE Department, S R Engineering College, Warangal, India

Kiran H Patil
Asst. Professor, CSE Department, S R Engineering College, Warangal, India

ABSTRACT

Have you ever desired to automatically wish your friends on their birthdays, or send a set of messages to your friend (or any WhatsApp connection!) automatically at a preset time, or send your friends by sending hundreds of random texts on WhatsApp and reply to your WhatsApp! messages automatically and in some cases, you are not next to your phone, but want to respond to WhatsApp messages? there is no way unless you stay online. Auto Responder for WhatsApp is an automatically respond to predefined messages, which contain few words or equal a message. Auto Responder WhatsApp is alternatively called WA Chat Bot. You can set custom responses for different messages. Each time a user enters a message, the library saves the text that they entered and the text that the message was in response to. As ChatterBot receives more input the number of responses that it can responds and the accuracy of each response in relation to the input statement increase. The code obtains precise response to the actual statement of search and results in appropriate response statement based on how frequently the person issues each response the bot communicates with. NLTK is written in Python and distributed under the GPL open source license. NLTK was the most promising technique in education and research from the past three decades. NLT is the most promising statistical and symbolic code module to process natural languages. User friendly interface being provided by NLTK to various libraries towards text processing viz., parsing, tokenization and streaming.

Key word: Natural Language Processing (NLP), Artificial Intelligence (AI), Chatter bot, Naïve Bayer's Classification, Jaccard Index.

Cite this Article: Y. Nagender, Kiran H Patil, WhatsApp Auto Responder using Natural Language Processing and AI. *International Journal of Computer Engineering & Technology*, 8(5), 2017, pp. 15–22.
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=8&IType=5

1. INTRODUCTION

Now a days WhatsApp is Trending and every person uses WhatsApp on earth, but we don't have WhatsApp Auto reply feature. WhatsApp auto replay is completely based on Python, ChatterBot and NLTK. Responds to your WhatsApp messages automatically, Requirements for auto responding, Python is a programming language that lets you work more quickly and integrate your systems more *effectively,* Chatter bot package is a machine-learning based conversational dialog engine build in Python which makes it possible to generate responses based on collections of known conversations. Chatter bot is trained to speak any natural language using language independent design. NLP is a natural language processing, computational linguistics, parsing, tagging, tokenizing, syntax, linguistics, language, natural language, text analytics and word net is based on the total maximum synset similarity between each word in each sentence.

2. NATURAL LANGUAGE TOOL KIT

NLT consists of vide range of codes and data sets wherein symbolic and NLP as built in subsets. Based on recent advancement the scenario of NLTK is as follows Consider the independent tasks of tagging and stemming, which both operate on sequences of tokens. If stemming is done first, we lose information required for tagging. If tagging is done first, the stemming must be able to skip over the tags. If both are done independently, we need to be able to align the results. As task combinations multiply, managing the data becomes extremely difficult. Pre-Processing the first step of pre-processing which is used to presents the string into clear word. The documentation to be prepared for text classification is categorized as follows. 1. Tokenization: A text is treated as a string, and then partitioned into a list of tokens? Removing stop words: Stop words such as “the”, “a”, “and”, etc. are frequently occurring, so the insignificant words need to be removed.? Stemming word: Applying the stemming algorithm that converts different word form into similar canonical form. This step is the process of conflating tokens to their root form, e.g. connection to connect, computing to compute. imports the tokenize package, defines a text string, and tokenizes the string on whitespace to create a list of tokens.

```
>>> from nltk. token import *
>>> Token (TEXT="Hello World!") <Hello World!>
```

In a similar fashion, other language processing tasks such as word-sense disambiguation, chunking and parsing all add properties to the Token data structure. In general, language processing tasks are formulated as annotations and transformations involving Tokens. Precisely, each processing task consider a token and extends it to include new information and this process is monotonic, i.e., to append the data only without addition or deletion of existing data. we can recognize ourselves that some words carry more meaning than other words. We can also see that some words are just plain useless, and are filler words. They may be used frequently in communicative “English” language such as “fluff”. The most regularly used phrase is “haa”. Even though it is not properly sentenced and grammatically incorrect, humans used to communicate with these meaning less words. For some responses we may use “ummm” or “uhha” and “uhh” which do not have any language base they do not find any space in database due to increased processing duration. These words are named as “stop words” since they cannot be processed. A different scenario of “stop words” is more literal i.e., “the words we stop on”.

2.1. Naïve Bayes Classifier

It is a classification technique based on Bayes theorem with an assumption of independence among predictors. It is considered to be the most suitable algorithm for text classification prior to training and testing. This would present you with some serious bias issues, so you should never train and test against the exact same data.

It is known to be supervised machine learning, since machine data is shown and, and coding it as "hey, this data is positive," or "this data is negative." After training new data and new data will be tested using the following code.

```
from nltk.corpus import movie_reviewsdocuments = [(list(movie_reviews.words(fileid)), category)a
            for category in movie_reviews.categories()
            for fileid in movie_reviews.fileids(category)]
```

Thereafter new classifier is defined as

```
Classifiera = nltk.NaiveBayesClassifier.train (training_set).
```

Initially the Naïve Bayes classifier is invoked and “train ()” is used to train the machine in one line.

2.2. Jaccard Index

The Jaccard index is considered as Inter section over Union and the Jaccard similarity coefficient. It is defined as the ratio of size of intersection and size of union of sample set based on finite sample set similarity which is measured by Jaccard index. In the numerator, is the count of items that are shared between the sets. The denominator is the total count of items across both the sets. Let's say we define sentences to be equivalent if 50% or more of their tokens are equivalent.

Here are two sample sentences:

The young cat is hungry.

The cat is very hungry.

The following sets results will be obtained by removing stop words,

{Young, cat, hungry}

{cat, very, hungry}

And the intersection is {cat, hungry}, with count of two.

The union of these sets is {young, cat, very, hungry}, with four word count.

Use a zero before decimal points: “0.25,” not “.25.” Use “cm^3,” not “cc.” Indicate sample dimensions as “0.1 cm × 0.2 cm,” not “0.1 × 0.2 cm^2.” The abbreviation for “seconds” is “s,” not “sec.” Use “Wb/m^2” or “Webbers per square meter,” not “webers/m^2.” When expressing a range of values, write “7 to 9” or “7-9,” not “7~9.”

A parenthetical statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.) In American English, periods and commas are within quotation marks, like “this period.” Other punctuation is “outside”! Avoid contractions; for example, write “do not” instead of “don’t.” The serial comma is preferred: “A, B, and C” instead of “A, B and C.”

If you wish, you may write in the first person singular or plural and use the active voice (“I observed that ...” or “We observed that ...” instead of “It was observed that ...”).

Remember to check spelling. If your native language is not English, please get a native English-speaking colleague to carefully proofread your paper.

2.3. Levenshtein Distance

The Levenshtein distance is a string metric for measuring the difference between two sequences. It is the most minimum distance between “single-character” edits viz., insertions, deletions or substitutions in order to vary two words. Levenshtein distance may also be referred to as edit distance.

1. The "kitten" and "sitting" has a Levenshtein distance of “3”, the following are the minimum possible variations

- kitten ? sitten (substitution of “s" for "k").
- sitten ? sittin (substitution of “i" for "e").
- sittin ? sitting (insertion of "g" at the end).

2. Compare two statements based on the Levenshtein distance of each statement's text. For example, there is a 65% similarity between the statements.

"Where is the post office?" and "looking for the post office” based on the Levenshtein distance algorithm.

3. OBJECTIVE AND METHODOLOGY

The objective of our work is to Design and Develop a Bot which will be able to answer questions. So, our Objective is to classify text data generated by WhatsApp user.

- ChatterBot is a Python library that makes it easy to generate automated responses to a user’s input. ChatterBot uses a selection of machine learning algorithms to produce different types of responses.
- NLTK is a broad-coverage natural language toolkit that provides a simple, extensible, uniform framework. NLTK is now widely used in research and teaching.

4. METHODOLOGY

We have developed architecture to extract response from text generated by WhatsApp user. Flow diagram of Methodology of our work is shown in figure.

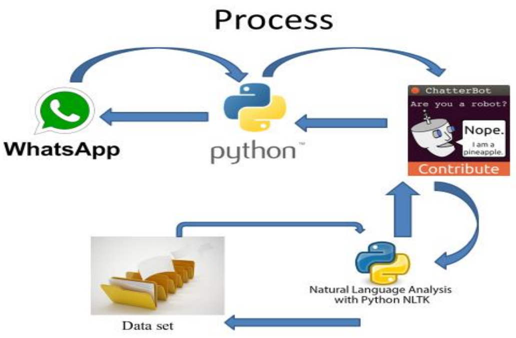


Figure 1 Proposed System Architecture

Chat Robot (ChatterBot) is a computer program to simulate human conversation, or chatting process, through AI technique. Usually, a Chatter bot is used to communicate with human beings.

“ChatterBot” is a built-in Python library which makes it easy to generate automatic response to user input. Chatterbot use multiple machine learning algorithms to generate different types of responses. It makes the task easier for developers to produce chat bots and automatic conversations with users. The flow diagram illustrate different ideas on Chatterbot’s.

An example of typical input would be something like this:

user: Good morning! How are you doing?

bot: I am doing very well, thank you for asking.

user: You're welcome.

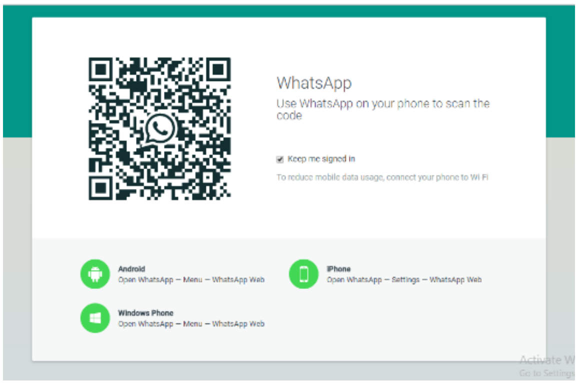
bot: Do you like hats?

```
from chatterbot import ChatBot

chatbot = ChatBot ('evreddy’,

trainer='chatterbot.trainers.ChatterBotCorpusTrainer’)

# Train based on the english corpus

chatbot.train("chatterbot.corpus.english")

# Print an example of getting one math based response

response = chatbot.get_response(“Hi")

print(response) # How are you?
```

Upon training with real conversations apart from regular sentence types, it is possible to generate appropriate and accurate conversation with Chatbot’s. it is based on case based reasoning instead of generalized data. All the data pertaining to past conversations will be stored and reply will be generated on explicit learning and an emotional and live content will be delivered by a chatbot.

5. RESULTS

It is concluded that a Chatbot use NLP and heuristic methods such as pattern matching and the functionality is more realistic compared to one approach method. A very general response may be achieved. These problems will be solved by statistic models by inducing a real emotional content and varied response. For instance, they may be well versed in discourse about their favorite music, but unable to say anything intelligible about a different type of music, simply because no such discourse was likely to be in the training corpora. My solution was to try to find a statistically relevant response, but revert to pattern matching responses if none could be found.



WhatsApp

Use WhatsApp on your phone to scan the code

☑ Keep me signed in

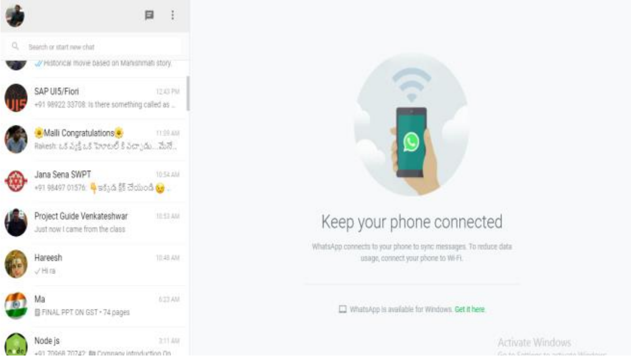
To reduce mobile data usage, connect your phone to Wi Fi

Android
Open WhatsApp – Menu – WhatsApp Web

iPhone
Open WhatsApp – Settings – WhatsApp Web

Windows Phone
Open WhatsApp – Menu – WhatsApp Web

Figure 2 WhatsApp Login Screen



Historical movie based on Mahishmati story.

SAP UI5/Fiori 12:43 PM
+91 98922 33708: Is there something called as ...

Malli Congratulations 11:09 AM
Rakesh: [illegible]

Jana Sena SWPT 10:54 AM
+91 98497 01576: 👇ఇక్కడ క్లిక్ చేయండి 😠 ...

Project Guide Venkateshwar 10:53 AM
Just now I came from the class

Hareesh 10:48 AM
✓ Hi ra

Ma 6:23 AM
FINAL PPT ON GST • 74 pages

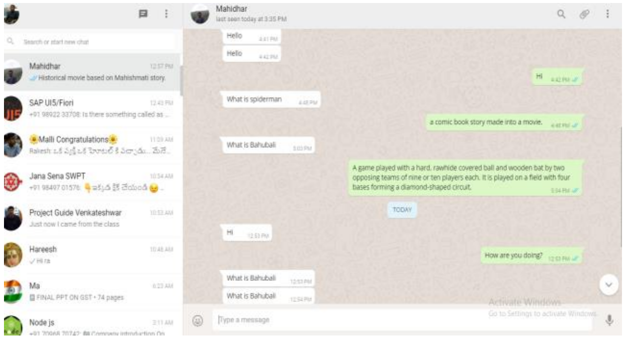
Node js 3:11 AM
+91 70968 70742: Company introduction On

Keep your phone connected

WhatsApp connects to your phone to sync messages. To reduce data usage, connect your phone to Wi-Fi.

WhatsApp is available for Windows. Get it here

Figure 3 WhatsApp Notification



Mahidhar 12:57 PM
Historical movie based on Mahishmati story.

Mahidhar
last seen today at 3:35 PM

Hello 4:41 PM

Hello 4:42 PM

Hi 4:42 PM

What is spiderman 4:48 PM

a comic book story made into a movie. 4:48 PM

What is Bahubali 5:03 PM

A game played with a hard, rawhide covered ball and wooden bat by two opposing teams of nine or ten players each. It is played on a field with four bases forming a diamond-shaped circuit. 5:34 PM

TODAY

Hi 12:53 PM

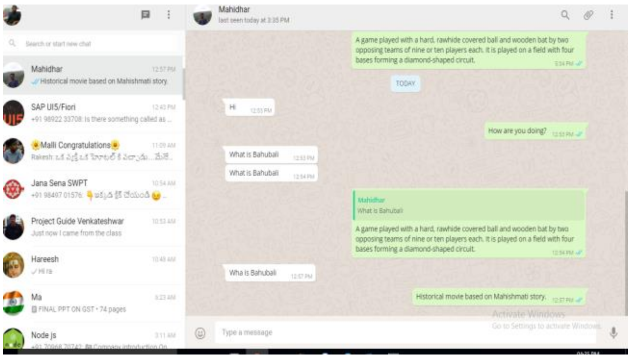
How are you doing? 12:53 PM

What is Bahubali 12:53 PM

What is Bahubali 12:54 PM

Type a message

Figure 4 WhatsApp Selected Chat



Mahidhar
last seen today at 3:35 PM

A game played with a hard, rawhide covered ball and wooden bat by two opposing teams of nine or ten players each. It is played on a field with four bases forming a diamond-shaped circuit. 5:34 PM

TODAY

Hi 12:53 PM

How are you doing? 12:53 PM

What is Bahubali 12:53 PM

What is Bahubali 12:54 PM

Mahidhar
What is Bahubali

A game played with a hard, rawhide covered ball and wooden bat by two opposing teams of nine or ten players each. It is played on a field with four bases forming a diamond-shaped circuit. 12:54 PM

Wha is Bahubali 12:57 PM

Historical movie based on Mahishmati story. 12:57 PM

Type a message

Figure 5 WhatsApp Response to Chat

6. CONCLUSIONS

We have presented a generally applicable for the auto responder for the WhatsApp chat messages by using a python Chatterbot. And Chatterbot has the capability of machine learning algorithms to produce different types of responses. This makes it easy for developers to create chat bots and automate conversations with users. The language independent design of ChatterBot allows it to be trained to speak any language. An untrained instance of ChatterBot starts off with no knowledge of how to communicate. Each time a user enters a statement, the library saves the text that they entered and the text that the statement was in response to. As ChatterBot receives more input the number of responses that it can reply and the accuracy of each response in relation to the input statement increase. The program selects the closest matching response by searching for the closest matching known statement that matches the input, it then returns the most likely response to that statement based on how frequently the people issues each response the bot communicates with. NLTK 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, it measures the semantic similarity of two concepts as their depth of the least common subsumer in the wordnet graph.

Future scope of WhatsApp Auto Responder is the ChatterBot should know the acknowledgement of message status. Clearly the results will improve with the quality and quantity of training data and the sophistication of heuristic rules. Therefore, over time I hope to add to the amount of statistical training data as well as develop a more comprehensive set of pattern matching rules. I also hope to extend memory capabilities throughout a session. Right now, my bots are limited to only remembering one username and which responses they have already used. It would be fascinating, however, to add the ability to learn more content during conversations. For example, a bot could learn to recognize sentences describing close relations to the speaker or what the speaker is doing now, and later ask about those people the speaker mentioned, or how the previous projects and activities went. The final improvement I would like to make is to make the application more scalable. Right now each bot is designed to handle talking to around one or two people at a time , and the conversations are not private. If I were to enable secure sessions and set browser cookies, I could run a distinct and private session for every user simultaneously. I hope to implement this whenever I obtain computing resources sufficient to handle more wide spread use.

REFERENCES

[1] J. Bang, H. Noh, Y. Kim and G. G. Lee, "Example-based chat oriented dialogue system with personalized long-term memory," 2015 International Conference on Big Data and Smart Computing (BIGCOMP), Jeju, 2015, pp. 256-265.

[2] C. Dewes, A. Wichmann, and A. Feldmann, “An analysis of Internet chat systems,” in Proc. ACM SIGCOMM IMC, Miami, FL, Oct. 2003, pp. 51–64.

[3] Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551

[4] Burstein, J. (2009). Opportunities for natural language processing research in education. In Computational Linguistics and Intelligent Text Processing (pp. 6-27). Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-00382-0_2

[5] Dr. Prasanth Yalla and Nakul Sharma, Combining Natural Language Processing and Software Engineering, In. Proc. International Conference in Recent Trends and Sciences

(ICRTES), March 14-15, 2014, Pages 370-373, ISBN: 978-93-5107-223-2. (Conference Proceedings)

[6] A. V. Yeole, P. V. Chavan, and M. C. Nikose, Opinion mining for emotions determination, ICIIECS 2015 - 2015 IEEE Int. Conf. Innov. Information, Embed. Commun. Syst., 2015.

[7] Bird, Klein, and Edward Loper.Natural Language Processing with Python. "O'Reilly Media, Inc.", 2009.

[8] Sebastiani F., "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[9] Miller G., "Nouns in WordNet: A Lexical Inheritance System", International Journal of Lexicography, vol. 3, no. 4, 1990.

[10] Kim S., Han K., Rim H., and Myaeng S. H. 2006. Some effective techniques for naïve bayes text classification. IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 11, pp. 1457-1466.

[11] Liu T., Chen Z., Zhang B., Ma W., and Wu G. 2004. Improving text classification using local latent semantic indexing. In proceedings of the 4th IEEE international conference on Data Mining, pp. 162-169.

[12] M. M. Saad Missen, and M. Boughanem. 2009. Using WordNet‟s semantic relations for opinion detection in blogs. ECIR 2009, LNCS 5478, pp. 729-733, Springer Verlag Berlin Heidelberg.

[13] Balahur A., and Montoyo A.. 2008. A feature dependent method for opinion mining and classification. In proceedings of the IEEE international conference on Natural Language Processing and Knowledge Engineering, pp. 1-7.

[14] Zhao L., and Li C.. 2009. Ontology based opinion mining for movie reviews. KSEM 2009, LNAI 5914, pp. 204-214, Springer-Verlag Berlin Heidelberg.

[15] Durant K. T., Smith M. D. 2006. Predicting the political sentiment of web log posts using supervised machine learning techniques coupled with feature selection, WebKDD 2006, LNAI 4811, pp. 187-206, Springer-Verlag Berlin Heidelberg.

[16] Polpinij J., and Ghose A. K. 2008. An ontology-based sentiment classification methodology for online consumer reviews. In proceedings of the IEEE international conference on Web Intelligence and Intelligent Agent Technology, pp. 518-524

[17] Deeptanshu Jha, Dr. Rashmi Jha, Varun Varshney, Natural Language Processing And Sanskrit, International Journal of Computer Engineering and Technology (IJCET), 5(10), 2014, pp. 57–63

[18] Natural Language Processing Based Information Retrieval For The Purpose Of Author Identification, International Journal of Information Technology & Management Information System (IJITMIS), 1(1), 2010, pp. 45–54

[19] NeeluNihalani, Dr. Mahesh Motwani, Dr. Sanjay Silakari, INtelligent Query Converter": A Domain Independent Interface For Conversion Of Natural Language Queries In English To Sql. International Journal of Computer Engineering and Technology (IJCET), 4(2), 2013, pp. 379–385