A Project Report on

COLLEGE ENQUIRY CHATBOT

by

Sayali Kamble

Vishal Jain

Zahid Khan

Under the Guidance of

Ms. Brinal Colaco

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road, Kasarvadavli, Thane(W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2018-2019

Approval Sheet

This Project Report entitled *"Pre-processing module of College Enquiry Chatbot"* Submitted by *"Sayali Kamble"(16102039),"Vishal Jain"(16102044),"Zahid Khan"(16102051)* is approved for the partial fulfillment of the requirement for the Mini Project .

Ms. Brinal Colaco
Guide

Prof. Sachin Malve
Head Department of Computer Engineering

Place :A. P. Shah Institute of Technology, Thane

Date:11/04/2019

CERTIFICATE

This is to certify that the project entitled *"College Enquiry Chatbot"* submitted by *"Sayali Kamble" (16102039),"Vishal Jain" (16102044),"Zahid Khan" (16102051)* for the partial fulfillment of the requirement for Mini Project is a bonafide work carried out during academic year 2018-2019.

Ms. Brinal Colaco
Guide

Prof. Sachin Malve
Head Department of Computer Engineering

External Examiner(s)

1.

2.

Place : A.P.Shah Institute of Technology, Thane

Date:11/04/2019

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified a idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Sayali Kamble , 16102039)

_____

(Vishal Jain , 16102044)

_____

(Zahid Khan ,16102051)

Date: 11/04/2019

Abstract

Chatbots are programs that mimic human conversation . It is designed to be the ultimate virtual assistant, entertainment purpose, helping one to complete tasks ranging from answering questions, getting driving directions, turning up the thermostat in smart home, to playing one's favorite tunes etc.

Chatbot has become more popular in business groups right now as they can reduce customer service cost and handles multiple users at a time. But yet to accomplish many tasks there is need to make chatbots as efficient as possible.

In this report dataset was collected and pre-processing was carried out on this dataset which cleans the dataset by removing the garbage values , replacing special characters with space , separating the dataset in question and answer format .

Contents

Keywords

Chatbots

Neural Networks

Machine learning

Sequence to Sequence modeling

Artificial intelligence

Natural language processing

TensorFlow

Chapter 1

Introduction

Todays era is having many web based services like E-business, Entertainment, Virtual assistance and many more. There is drastic increase in the world of web service, where everthing is now getting associated with web.

In this project, we will create a chatbot which will provide information about A. P. Shah Institute of Technology to the student through web interface using TensorFlow and NLP techniques. Thus, our bot will act as a person at enquiry office whom they can ask any questions regarding admission and college environment at any time in more interactive way. Thus, reducing workload of the admission cell personnel. This system can be integrated on college website and thus students who desires to get admission in the college can clear their doubts from their home. TensorFlow is used to create Neural model on which bot will be trained based on intent file. This model can be updated as and when new rule is implemented in the college.

Chapter 2

Literature Review

A. ***An Intelligent web-based voice chat bot***

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

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.

B. ***Chinese Intelligent Chat Robot Based on the AIML***

Wei Yun-Gang, Sun Bo, Sun Ming-Chen, Zhao Cui-Yi, and Ma PeiZi, "Chinese Intelligent Chat Robot Based on the AIML language", in Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, 2014, p. 368.

This paper explains the language enhancements in the field of Chatting Bot Development system. Here, the bot developed is in Chinese Language known as Chinese Intelligent Chat Robot Xiao Hui-hui.

This major technological enhancement in this research paper is language enhancement that AIML can be used not only to develop bots in English Language but also in many other foreign languages such as Chinese, Japanese, Indonesian, Hindi, Marathi etc.

C. *Development and Implementation of a chat bot in a Social Network*

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

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 parts 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:

1 Message reception.
2 Message processing.
3 Generation of a suitable reply.

Chapter 3

Data Pre-processing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. The dataset consists of some values that are null that is the conversation sentences ends half the way this type of conversations can't be provided for training the model . The pre-processing program will remove this garbage values then it will count the numbers of words in the dataset by word to count next it will distribute these words into two sets either question or answer this program also discards the words whose frequency will be less than 20 in the dataset so that is don't affect the duration of training end of Statement , Start of Statement , Out ,Pad are the defined variables used and indexed for defining the sentences .

Chapter 4

Result

Pre-processing of the dataset was completed . After the execution of the program the garbage values where nullified . one particular conversation is provided with unique number and every word in this conversation is provided with a index value . If the frequency of occurrence of a given word in the less then it gets the index value of out function in the code. The output consist of the sorted dataset according to the length of questions which will the helpful while fetching the values on the runtime conversations.

Chapter 5

Conclusions and Future Scope

Chatting bot service provider acts as a customer care for many organization /institutions / industries etc. or it may act as a personal assistant to all the people of the world. Main objective of this project was to provide a system to create chatbot which can be used by small businesses as a replacement of customer support. The further 3 steps of the project that is sequence to sequence modelling , training and testing will be done as our major project . Voice to text conversion can also be implemented for more user friendly experience of the Website .

Bibliography

1. B. R. Ranoliya, N. Raghuwanshi and S. Singh, "Chatbot for university related FAQs," in Advance in Computing, Communications and Informatics (ICACCI), 2017.
2. Chih-Wei Lee, Yau-Shian Wang, Tsung-Yuan Hsu, Kuan-Yu Chen, Hung-Yi Lee, Lin-shan Lee "scalable sentiment for sequence-to-sequence chatbot response with performance analysis" [ICASSP],2018.
3. N. Thomas, "An e-business chatbot using aiml and lsa," in Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on. IEEE, 2016, pp. 2740–2742.
4. A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao and B. Dolan, "A Neural Network Approach to Context-Sensitive Generation of Conversational Responses," in arXiv:1506.06714, 22 June 2015.S. O. Demokritov and V. E. Demidov, "Micro- Brillouin light scattering spectroscopy of magnetic nanostructures," IEEE Trans. Magn. , to be published.

Appendices

Appendix-A: Download and Installation

[1] https://www.python.org/downloads/

[2] https://www.anaconda.com/distribution/

Acknowledgement

We have great pleasure in presenting the report on **Pre-processing module of College Enquiry Chatbot .**We take this opportunity to express our sincere thanks towards our guide **Ms. Brinal Colaco** Department of Computer Engineering , APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof. Sachin Malve** Head of Department, Computer Engineering, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student Name1: Sayali Kamble

Student ID1: 16102039

Student Name2: Vishal Jain

Student ID2: 16102044

Student Name3: Zahid Khan

Student ID3: 16102051