Student Enquiry Chat-Bot System

Haritha.K, Dr.Asha.G

M.tech Computer Science and Engineering MES College of Engineering ,Kuttipuram Kerala, India
Assistant Professor,Computer Science and Engineering MES College of Engineering ,Kuttipuram Kerala,India

----

ABSTRACT—A Student bot project is built using artificial algo- rithms that analyzes users queries and understand users message. This System provides answer to the query of the student. Students just have to query through the bot which is used for chatting. Students can chat using any format there is no specific format the user has to follow. The System uses built in artificial intelligence to answer the query. The answers are appropriate what the user queries. The User can query any college related activities through the system. The user does not have to personally go to the college for enquiry. The System analyzes the question and then answers to the user. The system answers to the query as if it is answered by the person. With the help of artificial intelligence, the system answers the query asked by the students. The system replies using an effective Graphical user interface.This system helps the student to be updated about the college activities.

Index Terms—NLP, TextPreProcessing, TF-IDFApproach, Co- sine Similarity

I.   INTRODUCTION

A CHATBOT (also known as a talkbot, chatterbox, Bot, IM bot or Artificial Conversational Entity) is a computer program that mimics human conversations in its natural format including text or spoken language using artificial intelligence techniques such as Natural Language Processing (NLP), image and video processing, and audio analysis. It can replace a human for many tasks of answering queries. A chatbot is an agent that interacts with users using natural language. It was built as an attempt to fool humans. Several applications of chatbots such as Customer Service, call centers etc. uses Artificial Intelligence Markup Language to chat with user. One of the prime goals of chatbots is to resemble an intelligent human and make it difficult for the receiver of the conversation to understand the real working along with various architecture and capabilities for their usage has widely broadened. These chatbots can prove sufficient to fool the user into believing they are talking to a human being, but are very limited in improving their knowledge base at runtime, and have usually little to no means of keeping track of all the conversation data . Chatbots makes use of machine learning to reach artificial intelligence helping them to understand the user query and provide an appropriate response. The chatbots are developed using the Artificial Intelligence Markup Language for communicating or interacting with the user. This consist a software which will be made up using Artificial Intelligence and will help user to chat with machine.

Chat-bot for college management system project will be developed using artificial intelligence algorithms that will analyze users queries. This system will be a web application which will provide answers to the analyzed queries of the user. Artificial intelligence will be used to answer the user's queries. The user will get the appropriate answers to their queries.   The answers will be given using the artificial intelligence algorithms. Users won't have to go personally to the college for inquiry.

It is often impossible to get all the data on a single and simple interface without the complications of going through multiple forms and windows. The College Chatbot aims to re- move this difficulty by providing a common and user friendly interface that uses natural language processing to interact and solve queries of college students and teachers.

II.   LITERATURE SURVEY

A. State of the art
1) Content-Oriented User Modeling for Personalized Re- sponse Ranking in Chatbots : Automatic chatbots (also known as chat-agents) have attracted much attention from both researching and industrial elds. Generally, the semantic relevance between users queries and the corresponding re- sponses is considered as the essential element for conversation modeling in both generation and ranking based chat systems. By contrast, it is a

nontrivial task to adopt the usersinformation,such as preference,social role,etc.,into conversational models reasonably, while users proles play a signicant role in the procedure of conversations by providing the implicit contexts. Bingquan Liu , Zhen Xu , Chengjie Sun, Baoxun Wang

, Xiaolong Wang, and Derek F. Wong [1] are aims to address the personalized response ranking task by incorporating user proles in to the conversation model. In this approach, users personalized representations are latently learned from the contents posted by them via a two-branch neural network. After that, a deep neural network architecture is further presented to learn the fusion representation of posts, responses, and personal information. In this way, the proposed model could understand conversations from the users perspective; hence, the more appropriate responses are selected for a specied person.The experimental results on two datasets from social network services demonstrate that our approach is hopeful to represent users personal information implicitly based on user generated contents,and it is promising to perform as an

important component in chatbots to select the personalized responses for each user.

2) Designing a Chat-bot that Simulates an Historical Fig- ure : There are many applications that are incorporating a human appearance and intending to simulate human dialogue, but in most of the cases the knowledge of the conversational bot is stored in a database created by a human experts. How- ever, very few researches have investigated the idea of creating a chat-bot with an artificial character and personality starting from web pages or plain text about a certain person. Emanuela Haller and Traian Rebedea [2] describes an approach to the idea of identifying the most important facts in texts describing the life (including the personality) of an historical figure for building a conversational agent that could be used in middle- school CSCL scenarios. In here first Extracting information from Wikipedia. Then Designing the conversation. Chat script is a scripting language which is designed to accept user text input and generate a text response.

3) Keyword-based Search and Exploration on Databases : Empowering users to access databases using simple keywords can relieve users from the steep learning curve of mastering a structured query language and understanding complex and possibly fast-evolving data schemas. Here Yi

Chen,WeiWang and Ziyang Liu [3] give an overview of the state-of-the-art techniques for supporting keyword-based search and exploration on databases. Several topics will be discussed, including query result denition, ranking functions, result generation and top-k query processing, snippet generation, result clustering, result comparison, query cleaning and suggestion, perfor- mance optimization, and search quality evaluation. Various data models will be discussed, including relational data, XML data, graph-structured data, data streams, and workows.

4) chatbot using A knowledge in Database: Human-to- Machine Conversation modeling: A chatterbot or chatbot aims to make a conversation between both human and machine. The machine has been embedded knowledge to identify the sentences and making a decision itself as response to answer a question. Bayu Setiaji and Ferry Wahyu Wibowo are proposed

[4] a chatbot using knowledge in database.In here the response principle is matching the input sentence from user. From input sentence, it will be scored to get the similarity of sentences, the higher score obtained the more similar of reference sentences. The sentence similarity calculation using bigram which divides input sentence as two letters of input sentence. The knowledge of chatbot are stored in the database. The chatbot consists   of core and interface that is accessing that core in relational   database   management   systems (RDBMS). The database has been employed as knowledge storage and interpreter has been employed as stored programs of function and procedure sets for pattern-matching requirement.

5) Example-based chat oriented dialogue system with per- sonalized long- term memory: J. Bang, H. Noh, Y. Kim and

G. G. Lee [5] introduces an example-based chatoriented dia- logue system with personalization framework using long-term memory. Previous representative chat-bots use simple keyword and pattern matching methodologies. To maintain the qualityof systems, generating numerous heuristic rules with human labour is inevitable. The language expert knowledge is also necessary to build those rules and matching patterns. To avoid high annotation cost, example-based dialogue management is adopted for building chat-oriented dialogue system. Here also propose three features: POS-tagged tokens for sentence match- ing, using NE types and values for searching proper responses, and using back-off responses for unmatched user utterances.

Also, this system automatically collects user-related facts from user input sentences and stores the facts into a long-term memory. System responses can be modied by applying user- related facts in the long-term memory. A relevance score of a system response is proposed to select responses that include user-related fact, or frequently used responses. In several experiments, have found that this proposed features contribute to improve the performance and this system shows competitive performance to ALICE system with the same training corpus.

III. PROPOSED METHOD

Python is a software that provides user-friendly interface   to make the connection easier and convenient with the in- ternet providing valid and reliable web services.Here created a sample chatbot using the same with twitch as an online platform which provides a chatbot platform to the clients who are online. Web based platform provides a vast   intelligent base which can be helpful in simulating problem solving for humans. proposed system Focusing the usage of chatbots in the field of education. A Student bot project is built using artificial algorithms that analyzes users queries and understand users message. Students just have to query through the bot which   is used for chatting. The answers are appropriate what the user queries. If the answer found to invalid, user just need to select the invalid answer button which will notify the admin about the incorrect answer. Admin can view invalid answer through portal via login. System allows admin to delete the invalid answer or to add a specific answer of that equivalent question. With the help of artificial intelligence, the system answers the query asked by the students. The system replies using an effective Graphical user interface which implies that as if a real person is talking to the user.

A. Phases involved in this Bot
There are some Phases involved in making of this bot

1) Get the User Query. :
2) Text Pre- Processing with NLTK: The main issue with text data is that it is all in text format (strings). However, Machine learning algorithms need some sort of numerical feature vector in order to perform the task. So before we start with any NLP project we need to pre-process it to make it ideal for work. Basic text pre-processing includes:
- Converting the entire text into uppercase or lowercase,   so that the algorithm does not treat the same words in different cases as different.

- Tokenization: Tokenization is just the term used to de- scribe the process of converting the normal text strings
into a list of tokens i.e words that we actually want. Sen- tence tokenizer can be used to find the list of sentences and Word tokenizer can be used to find the list of words in strings.The NLTK data package includes a pre-trained Punkt tokenizer for English.
- Removing Noise: i.e everything that isnt in a standard number or letter.
- Removing Stop words: Sometimes, some extremely com- mon words which would appear to be of little value in helping select documents matching a user need are excluded from the vocabulary entirely. These words are called stop words.
- Stemming: Stemming is the process of reducing inflected (or sometimes derived) words to their stem, base or root form generally a written word
- form. Example if we were to stem the following words: Stems, Stemming, Stemmed, and Stemtization, the result would be a single word stem.
- Lemmatization: A slight variant of stemming is lemma- tization. The major difference between these is, that, stemming can often create non-existent words, whereas lemmas are actual words. So, your root stem, meaning the word you end up with, is not something you can just look up in a dictionary, but you can look up a lemma. Examples of Lemmatization are that run is a base form for words like running or ran or that the word better and good are in the same lemma so they are considered the same.
3) After the initial preprocessing phase, we need to trans- form the text into a meaningful vector (or array) of numbers. The bag-of-words is a representation of text that describes the occurrence of words within a document. It involves two things:
:
- A vocabulary of known words.
- A measure of the presence of known words.

Why is it is called a bag of words? That is because any information about the order or structure of words in the document is discarded and the model is only concerned with whether the known words occur in the document, not where they occur in the document.The intuition behind the Bag of Words is that documents are similar if they have similar content. Also, learn something about the meaning of the document from its content alone. For example, if the dictionary contains the words

Learning, is, the, not, great, and want to vectorize the text Learning is great, here would have the following vector: (1, 1, 0, 0, 1).

TF-IDF Approach

A problem with the Bag of Words approach is that highly frequent words start to dominate in the document (e.g. larger score), but may not contain as much informational content. Also, it will give more weight to longer documents than shorter documents.

One approach is to rescale the frequency of words by how often they appear in all documents so that the scores

for frequent words like the that are also frequent across all documents are penalized. This approach to scoring is called Term Frequency-Inverse Document Frequency, or TF-IDF for short, where:

Term Frequency: is a scoring of the frequency of the word in the current document.

TF = (Number of times term t appears in a document)/(Number of terms in the document)

Inverse Document Frequency: is a scoring of how rare the word is across documents.

$$IDF = 1 + \log(N/n) \qquad (1)$$

where, N is the number of documents and n is the number of documents a term t has appeared in.

Tf-IDF weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus.

4) Cosine Similarity. : TF-IDF is a transformation applied to texts to get two real-valued vectors in vector space. Then obtain the Cosine similarity of any pair of vectors by taking their dot product and dividing that by the product of their norms. That yields the cosine of the angle between the vectors. Cosine similarity is a measure of similarity between two non-zero vectors. Using this formula ,find out the similarity between any two documents d1 and d2.

$$CosineSimilarity(d1, d2) = Dotproduct(d1, d2)/\|d1\|\|d2\| \qquad (2)$$

where d1,d2 are two non zero vectors.

This student bot project built using Django framework.Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so focus on writing app without needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation, and

many options for free and paid-for support.]If match with the knowledge base then provide an appropriate response.

This student bot project built using Django frame- work.Django is a high-level Python web framework that en- ables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so focus on writing app with- out needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation, and many options for free and paid-for support.

IV. EXPERIMENTS AND RESULTS

The proposed system will be used to identify answers related to user submitted questions. The need is to develop a database where all the related data will be stored and to develop a web interface. A database is developing, which store information about questions, answers, keywords, logs and feedback messages. Given an input Question and generates a natural language answer as the output. The interactive student enquiry chatbot will take the query from the user and will give the appropriate answer to the user query. Here used a pre-trained artificial intelligence module so that it use its pretrained neural networks to answer the users query with efficiency and accuracy. Here also made some custom modules or entities such as calendar, time-table etc.to make it ample to answer college related queries.Emanuela Haller and Traian Rebedea "Designing a Chat-bot that Sim- ulates an Historical Figure", Conference Publications, July 2013.

[2] Yi Chen,WeiWang,Ziyang Liu "Keyword- based Search and Exploration on Databases", IEEE 27th International Conference on Data Engineer- ing, may 2011 .

[3] Bayu Setiaji, Ferry Wahyu Wibowo "Chatbot Using A Knowledge in Database Human-to-Machine Conversation Modeling", 2016 7th Inter- national Conference on Intelligent Systems, Modelling and Simulation.

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

[5] Bhavika R. Ranoliya, Nidhi Raghuwanshi and Sanjay Singh "Chatbot for University Related FAQs.", 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) .

Fig. 1. Result after Testing Dataset

Fig. 2. student chatbot output

V. CONCLUSION

This is a chatbot for educational sector. The main objectives of the project were to identify answers related to user submit- ted questions. User does not have to go personally to college office for the enquiry.

REFERENCES

[1]. Bingquan Liu , Zhen Xu , Chengjie Sun, Baoxun Wang , Xiaolong Wang,Derek F. Wong and Min Zhang, " Content-Oriented User Mod- eling for Personalized Response Ranking in Chatbots", IEEE/ACM transactions on audio, speech, and language processing, vol. 26, no. 1, january 2018 .