EasyChair Preprint

№ 3303

Chatbot using Data Sciences for Educational Institutions

Madhurja Nayan Deka, Kuntimaddi Ajay Kumar and A Christy

April 30, 2020

ChatBot using Data Sciences for Educational Institutions

Madhurja Nayan Deka[1] ,Kuntimaddi Ajay Kumar[2],A.Christy[3]

[1,2]CSE Dept, Sathyabama Institute of Science & Technology, Chennai, India

[3]Professor, CSE Dept, Sathyabama Institute of Science & Technology, Chennai, India

Abstract: In today's world computers play a crucial role in our society. Computers offer information; they entertain and facilitate in immeasurable manners. A chatbot may be a program designed to counterfeit a wise communication on a text or spoken ground. However, this paper relies on the text solely chatbot. Chatbot acknowledge the user input still as by mistreatment pattern matching, access data to produce a predefined acknowledgment supported the sentence given by the user. Once the input is being put into being within the information, a response from a predefined pattern is given to the user. A Chatbot is enforced mistreatment pattern scrutiny, within which the order of the sentence is recognized, and a saved response pattern is acclimate to the exclusive variables of the sentence. they can't register and reply to complicated queries, and area unit unable to perform compound activities. Chatbot is comparatively a replacement technology. the appliance of a Chatbot are often seen in numerous fields within the future. This paper covers the techniques accustomed style and implementation of a reliable Chatbot in terms of Educational Institutions.

KEYWORDS : Chatbot, DataScience, Educational, Similarity, Euclidean, etc.

INTRODUCTION

Dialogue systems are often divided into goal-driven systems like technical support services, and non-goal-driven systems, like learning tools or game characters. ancient dialogue systems is slot- fillingthat predefines the structure of a dialogue state as a group of slots to be crammed throughout the dialog. associate degree existing implementation may be a eating house reservation system, wherever the slots are often location, worth vary or sort of cooking of a eating house. Slot-filling has well-tried to be reliable, however laborious to scale to new domains. Slot filling uses rule primarily based approach eg., siri. during a rule primarily based approach, a larva answers question supported some rules on that it's trained on. The creation of those bots area unit comparatively simple mistreatment some rule-based approach, however the larva isn't economical in respondent queries, whose pattern doesn't match with the foundations on that the larva is trained. Those bots are often created by mistreatment language like computer science Markup Language(AIML), a language supported XML that enable developer's write rules for the larva to follow. Another disadvantage is writing rules for various situations is incredibly time intense and it's attainable} to put in writing rules for each possible state of affairs. therefore these bots will handle easy queries however fail to manage complicated queries.

Proposed system is making bots that use Machine Learning-based approach that create them a lot of economical than rule-based bots. These bots are called Retrieval primarily based models. These bots area unit trained on a group of queries and their potential outcomes. for each question, the larva will realize the foremost relevant answers from the sets of all potential answers then outputs the solution. Although, the larva cannot generate new answers if trained on

heaps of question and answer dataset, and if the info set is pre-processed neatly, the larva will handle queries fairly sensible. The quality will vary from easy rules for a question to complicated rules mistreatment some machine learning formula to search out the foremost acceptable answer. Also, there's no issue with the language and synchronic linguistics because the answers area unit pre-determined and it cannot fail in syntax manner. Our system uses machine learning model known as memory networks. Memory network uses logical thinking elements combined with an extended term memory element and together it's used for predicting the responses. Memory network is trained finish to end and therefore needs less direction throughout coaching, so it are often used for realistic settings. All the elements of the end-to-end systems area unit trained on education institution based queries & dialogues, but it is flexible enough to be implemented in other fields as well.

LITERATURE REVIEW

A. "Intelligent Chatbot for radio-controlled Navigation of Repository Contend", Miss Anjali Mishra ; Miss Shruti Sapre ; Miss Shruti Shinde ; Miss Shreya Nahar ; faculty member. S.N Shelke

The device projected here is Associate in Nursing interactive Application, that is capable of responsive a queries and Answer. we have a tendency to propose to develop interactive academic software package which might run on the desktop. The software package helps the user to induce answer while not reading the file. ab initio the software package is given input with the file of specific format . Most of the operating individuals in Asian nation by victimisation this software package can save time and find answer on click. In Future Scope Chat larva application will be build for numerous|thevaried|the assorted} completely different fields in several domains along side various domain. For Ex, For communication for Medical, Collages and different vital purpose.

B. "An Intelligent Behaviour shown by a ChatBot System", Vibhor Sharma ; Monika Goyal ; Drishti lead

A system which is able to work as Associate in Nursing application and provides users info concerning completely different styles of university connected queries. This application can work employing a pattern matching algorithmic program victimisation depth initial search (DFS). during this project, our responsibilities enclosed reading the user inputs so reply to the question, whereas making an attempt to stay the oral communication associated with University surroundings. the primary step in developing the FAQ larva consisted of intensive group action and writing down as several queries as doable. This power-assisted in permitting FAQ larva to showing intelligence match pattern (inputs). For doing that we have a tendency to created new AIML files and matched it with the colloquial cognitive content of ALICE larva

C. "Keyword Weighting Function for Document Clustering", A.Christy ; G.M.Gandhi ; S.Vaithyasubramaniam

Christy et al (2019) has proposed a keyword weighting function for document clustering. Each keyword in the sample are clustered based on keyword weighting function. Experimental results were conducted with BBC news collection related to 5 domains and compared with K-Means clustering and Hierarchical clustering algorithms. It is shown that clustering followed by keyword weighting function has improved accuracy [14].

D. "Prediction Based Interactive Application for Medical Field", M. Gandhi ; V. K. Singh ; V. Kumar

M. Gandhi, V. K. Singh and V. Kumar describes an interactive application analyzes symptoms to diagnose, predict medical conditions, generates treatments and suggestions based on the inputs provided by the user. In addition to that, the app tracks user's health activities like their step counts, sleep tracking, heart rate sensing and other parameters and displays users their periodic health reports. It incorporates various fitness activities tracked and other factors like their

age, gender, location, past medical records, and calories intake to perform a more accurate analysis [15].

ARCHITECTURE

This Educational Chatbot is incredibly easy and user friendly. it's not terribly sophisticated like alternative Chatbots. The operating of the Chatbot is straightforward and may be simply understood by somebody. In most cases, the operating is incredibly sophisticated. several categories area unit used that is tough to know. But during this program implementation, just one category is employed to create it easy and procure the expected output. This Educational Chatbot uses easy pattern matching to represent the input and output whereas alternative Chatbots uses input rules, keyword patterns and output rules to come up with a response. If the input isn't found within the information, a default response is generated. The input and output are often bespoke in step with the user. supported the developer or the user, the specified requests and responses are often hold on within the information. Since own information are often created, it permits the user to know however the response is generated. Chatbot may be a laptop application that uses computer science to mimic human spoken communication. It helps the user by respondent the queries asked by them. The program is enforced mistreatment Java programing language**.**

The following facts square measure unbroken in mind throughout planning a Chatbot :

A. Choice of OS

Windows is employed for this project as a result of it's user friendly. it's additionally strong.

B. Choice of Package

Eclipse package is employed for programming in java. as a result of it contains basic space and it's largely used for java applications

C. Making a Chatbot

For making a Chatbot, a program has got to be written. Java programing language is employed for programming. The Chatbot is formed in such how to assist the user, improve the communication and amuse the user.

D. Making a Talk

The chat is formed employing a pattern that's famed to the user and will be straightforward to know. Chat window show up to form spoken language. This window is formed exploitation java applets.

E. Pattern Matching

It is a method of AI utilized in the look of a Chatbot. The input is matched with the inputs saved within the info and corresponding response is came.

F. Simple

The design of a Chatbot is incredibly straightforward. It simply answers to the queries asked by the user, if the question is found within the info.

G. Colloquial and Amusive

The Chatbot responses square measure how famed to the user. The spoken language follows a Basic English language and interacts in a straightforward to browse manner. The spoken language between the user and therefore the larva is amusive. it's like lecture alternative person.

The whole process is distributed into various models as depicted in Fig. 1:

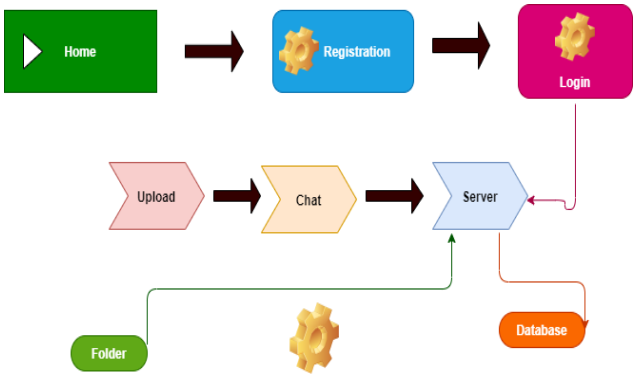

Fig.1 System Architecture

A. Registration Module

It handles the initialization of the User Interface between the user & the Chatbot.It consists of the details like Name, Registration

Number, Department, Year.

B. Login Module

It handles the user credentials required to access the Chatbot& verifies the input with registered details.

C. Bot Module

It involves the Bot & the User interacting with each other in a chat window.Here the user can ask questions either based on common activities or academic related.

D. Admin Module

It is where the Administrator will be adding Academic details & answers to common user queries.It also enables the management of registration & login information as shown in Fig.2

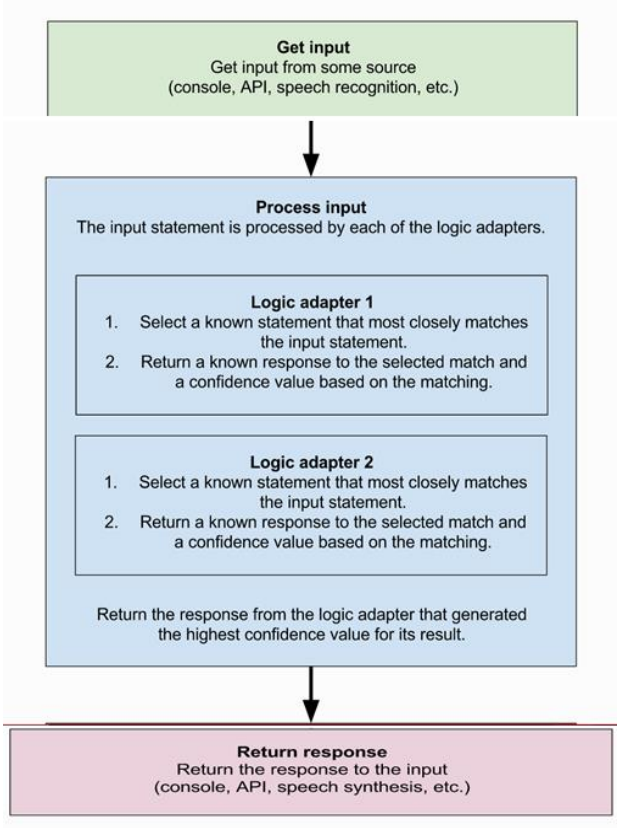

Fig.2 Working Modules

DATA SCIENCE IN CHATBOT

We implement Data Science based corpus modules to simplify the answering process. For extracting the information, we implement both single layer & double layer similarity based methods namely String Based & Text Based. This involves training our Bot in various data sets to get the relative queries perfectly answered to the user as shown in Fig.3

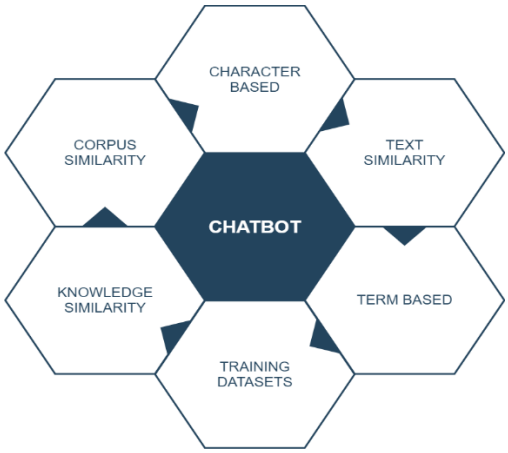

Fig.3 Role of Data Science

A. Cosine Similarity

Classical approach from linguistics is to live similarity supported the content overlap between documents. For this we'll represent documents as bag-of-words, thus every document are a distributed vector. And outline live of overlap as angle between vectors in defined in eqn. (1)

similarity(doc1,doc2)=cos(θ)=doc1doc2|doc1||doc2| ---------- (1)

By circular function distance/dissimilarity we tend to assume eqn(2)

distance(doc1,doc2)=1−similarity(doc1,doc2) ……..(2)

B. Euclidean Distance

Euclidean distance isn't thus helpful in IP field as Jaccard or trigonometric function similarities. however it forever value to undertake completely different measures. In text2vec it will by computed solely on dense matrices.

RESULTS AND DISCUSSION

This Educational Chatbot is incredibly straightforward and user friendly. it's not terribly sophisticated like alternative Chatbots.

The operating of the Chatbot is easy and may be simply understood by any individual. In alternative Chatbots, the operating is incredibly sophisticated. several categories square measure used that is tough to know. during this program, just one category is employed to create it straightforward and acquire the expected output. This Chatbot uses straightforward pattern matching to represent the input and output whereas alternative Chatbots uses input rules, keyword patterns and output rules to get a response. If the input isn't found within the info, a default response is generated. The input and output may vary with each & every individual user. supported the developer or the user, the specified requests and responses may be keep within the info. This Chatbot may be used for the Education Purpose & is also flexible enough to be implemented in other fields. Whenever an individual tends to have doubts or queries, he can reach out to the Bot for immediate & reliable response as shown in Fig.4. This is extremely helpful for the student as well as the institution and moreover saves a lot of time & effort.

Fig.4 Sample Chatbot

CONCLUSION

A chatbot is in a very one amongst one in every of the straightforward ways in which to move knowledge from a laptop while not having to suppose for correct keywords to seem up in a search or browse many websites to gather info; users will simply kind their question in linguistic communication and retrieve information. during this paper, info concerning the look, implementation of the chatbot has been given. From the survey higher than, it may be aforementioned that the event and improvement of chatbot style grow at hit or miss rate because of style of strategies and approaches wont to style a chatbot. They assist North American country by providing amusement, saving time and responsive the queries that square measure exhausting to search out. The Chatbot should be straightforward and colloquial. Since there square measure several styles and approaches for making a chatbot, it may be at odds with business concerns. Researchers have to be compelled to act and should agree on a typical approach for planning a Chatbot. during this project, we have a tendency to looked into however Chatbots square measure developed and therefore the applications of Chatbots in numerous fields. additionally comparison has been created with alternative Chatbots. General purpose Chatbot should be straightforward, user friendly, should be simply understood and therefore the cognitive content should be compact. though a number of the business product have recently emerged, enhancements should be created to search out a typical approach for planning a Chatbot.

REFERENCES

[1] R. S. Russell, "*Language Use, Personality and True Conversational Interfaces*", Project Report of AI and CS-University of Edinburgh, Edinburgh, pp.1-80,2002.

[2] Y. Zhou, X. Ziyu, A. W. Black, A. I. Rudnicky, "*Chatbot Evaluation and Database Expansion via Crowdsourcing*", Proc. of the Chatbot Workshop of LREC, US, pp. 16-19, 2016.

[3] C. R. Anik, C. Jacob, A. Mohanan, "*A Survey on Web Based Conversational Bot Design*", JETIR, Vol.3, Issue.10, pp.

96-99, 2016.

[4] R. P. Schumaker, H. Chen, "*Leveraging Question Answer Technology to Address Terrorism Inquiry*", Decision Support Systems, Vol.4, Issue.3, pp. 1419-1430,2007.

[5] B. P. Kiptonui, "*Chatbot Technology: A Possible Means of Unlocking Student Potential to Learn How to Learn, Educational Research*", Vol.4, Issue.2, pp. 218-221,2013.

[6] S. Ghose, J. J. Barua, "*Toward the Implementation of a Topic Specific Dialogue Based Natural Language Chatbot as an Undergraduate Advisor*", International Conference on Informatics, Electronics & Vision, India, pp. 1-5,2013.

[7] R. S. Russell, "*Language Use, Personality and True Conversational Interfaces*", Project Report of AI and CS-University of Edinburgh, Edinburgh, pp.1-80,2002.

[8] Y. Zhou, X. Ziyu, A. W. Black, A. I. Rudnicky, "*Chatbot Evaluation and Database Expansion via Crowdsourcing*", Proc. of the Chatbot Workshop of LREC, US, pp. 16-19, 2016.

[9] C. R. Anik, C. Jacob, A. Mohanan, "*A Survey on Web Based Conversational Bot Design*", JETIR, Vol.3, Issue.10, pp. 96-99, 2016.

[10] R. P. Schumaker, H. Chen, "*Leveraging Question Answer Technology to Address Terrorism Inquiry*", Decision Support Systems, Vol.4, Issue.3, pp. 1419-1430,2007.

[11] B. P. Kiptonui, "*Chatbot Technology: A Possible Means of Unlocking Student Potential to Learn How to Learn, Educational Research*", Vol.4, Issue.2, pp. 218-221,2013.

[12] S. Ghose, J. J. Barua, "*Toward the Implementation of a Topic Specific Dialogue Based Natural Language Chatbot as an Undergraduate Advisor*", International Conference on Informatics, Electronics & Vision, India, pp. 1-5,2013.

[13] J. Jia, "*The Study of the Application of a Keywords-based Chatbot System on the Teaching of Foreign Languages*", Report of University of Augsburg, Augsburg, , pp.1-36,2003.

[14] Christy, A., Gandhi, G.M., Vaithyasubramanian, S. (2019)," Clustering of text documents with keyword weighting function", International Journal of Intelligent Enterprise, Vol. 6, Issue.1, DOI: 10.1504/IJIE.2019.100029

[15] M. Gandhi, V. K. Singh and V. Kumar, "IntelliDoctor - AI based Medical Assistant," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019, pp. 162-168.

[16] S.Prayla shyry, Asha Deepika,R.Subhashini"Efficient identification of bots by K-means clustering " Proceedings of the International Conference on Soft Computing Systems (pp.307-318)",2016

[17] Jesudoss A. and Subramaniam N.P., "Enhanced Kerberos Authentication for Distributed Environment", Journal of Theoretical and Applied Information Technology, 2014 Vol. 69, No. 2, pp. 368-374.

[18] Joseph Manoj, R., Anto Praveena, M.D., Anvesh, M., Pujith, M. "Secured User Behaviour Based Access Framework for Web Service", IOP Conference Series: Materials Science and Engineering, 2019, Vol.590, pp.1-12.

[19] M.S.Roobini,DrM.Lakshmi,(2019),"Classification of Diabetes Mellitus using Soft Computing and Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering,ISSN: 2278-3075, Volume-8, Issue-6S4