International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

AI and Web-Based Interactive College Enquiry Chatbot

Akshada Phalle[1], Saniya Kadam[1], Sakshi Sonphule[1], Ila Savant[2]

[1]Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Maharashtra, India

---

Abstract – A chatbot is often described as the most advanced and promising expression of interaction between humans and machines. These digital assistants streamline the interactions between people and services with less human interaction. This is proved as to be a boon during the COVID-19 pandemic when people could no longer visit information desks to get their queries answered. Admission season is arguably the busiest time for a college office. Concerned parents flock the enquiry desks with inquiries about the college. These questions are repetitive and one can use a virtual agent for these doubts which would free the office personnel to handle other issues. This paper proposes a college enquiry chatbot based on Python's chatbot framework, Rasa. Rasa is an open source chatbot framework based on machine learning. With its help, one can easily create highly accurate chatbots and integrate them with websites and messaging applications such as WhatsApp, Facebook, Telegram, etc. The proposed chatbot could handle manually typed as well as rich responses generated by clicking on a payload. It could be integrated with the college website by adding a widget and deployed as on the messaging application, Telegram.

Keywords: Conversational AI, Natural Language Processing, Human Computer Interaction, Chatbot, College Enquiry Chatbot, Rasa, Rasa X, Python, Telegram Chatbot

1. INTRODUCTION

Most websites today be it real-estate, educational institutions, financial firms have a dedicated page for contact information in case a visitor has questions. The problem with this traditional form of online communication is that visitors have to contact help desks only within a stipulated time even if the query is trivial. In rush hours, most people have to be kept on hold in case of limited human resources. Although this is unavoidable, it may leave clients dissatisfied and sully a company's image. This is where conversational AI comes into the picture in the form of chatbots. An artificial conversation entity called chatbot system uses conversational Artificial Intelligence (AI) to simulate a dialogue. It uses rule-based language applications to perform live chat functions in response to real –time user interactions. It removes barriers to customer support that can occur when demand outpaces resources. Instead of waiting on hold, an enquirer can get answers to his/her questions in real time. This virtual assistant could provide valuable help to college institutions, especially during the admission season. In most colleges, faculties have to devote their time in the admission cells to answer the questions of concerned parents about a college. These queries, though important, are mostly redundant. Thus, we have proposed a web-based college enquiry chatbot that could be integrated with the college's website that could handle all these questions from various departments to placement cells to hostels and so on at a single click from a user. It allows the users to type their query manually or to click on a button of their choice in case they are unsure about their actual doubt. The system would contain information from the official website, except the user won't have to navigate to multiple pages just to know the cut-off for a particular department. Furthermore, for those who do not have the time to actually visit the college website, we have also proposed integration with the messaging application, Telegram which provides users with the opportunity to look back to their previous conversations with the bot, an addition to the web version.

2. MATHEMATICAL MODEL

Rasa NLU internally uses the following two algorithms for entity extraction:

1: Bag of words (BoW) is one of the famous methods for extracting the characteristics from conversations so that it can be used in building models for a machine learning algorithm. It is a representation of words which includes two main things, a document of known words and the frequency of known words. It can simply be simple or complex depending upon the complexity of creating the vocabulary of known words and the circumstance of known words. It follows the succeeding steps

a): Collecting the data which could be any input data taken from the user, we can treat each and every line as a different document.

b): Designing the vocabulary where we gather the list of all the unique words and discarding case sensitive and punctuation and put that into the model vocabulary.

c): Creating the document vector where we check the frequency of words in each document. The main thing of this step is to make sure that it turns each word in the document into the vector form like a matrix so that it can be easily used as an input to the other machine learning algorithms. The easiest way of scoring is to mark the presence of words as a Boolean value, 0 for absent and 1for present and then convert it into a binary vector.

2: Conditional Random Fields (CRFs) are a discriminative model, used for predicting sequences. They use contextual information from previous labels, thereby increasing the amount of information the model has, to make a good prediction.

We used the following parameters to evaluate the performance of our chatbot at each stage.

1. Intent accuracy= $\frac{number\ of\ correctly\ recoginzed\ intents}{total\ number\ of\ intents}$

2. Entity extract accuracy = $\frac{number\ of\ correctly\ extracted\ entittes}{total\ noumber\ of\ entites}$

3. Integrity of an entity = $\frac{number\ of\ reconginzed\ entites}{total\ number\ of\ entites}$

4. Integrity of a sentence = $\frac{number\ of\ cmpletely\ extracted\ lines}{total\ number\ of\ lines}$

3. METHODOLOGY

A Rasa NLU pipeline usually incorporates three main parts:

1. Tokenization: Tokenization involves breaking the raw text into small chunks called tokens. These tokens help in comprehending the context or developing a model for the NLP. It helps in explicating the meaning of the text by analyzing the progression of the words. There are two tokenizers which we have used in our chatbot.

- Whitespace Tokenizer: A Whitespace Tokenizer is a tokenizer that splits on and discards only whitespace characters. This implementation can return Word, Core Label or other LexedToken objects. It also has a method for whether to make EOL a token or whether to treat EOL characters or as a whitespace.

- SpacyTokenizer: It processes the text from left to right. First, the tokenizer splits the text on whitespace similar to the split() function in Python. Then it checks whether the substring matches the tokenizer exception rules.

2. Featurization: Featurization is the process to transfigure varied forms of data into numerical data. This data can be in the form of text, images, videos, graphs, various database tables, time-series, categorical features, etc. The featurizers used in this project are as follows:

- LanguageModelFeaturizer: This featurizer that uses transformer-based language models.It uses a pre-trained language model to compute vector representations of input text.

- RegexFeaturizer:It creates a vector representation of a user message using regular expressions. During training, the RegexFeaturizer will make a list of regular expressions as per the training data format. For each regex, a parameter will be set marking whether this expression was found in the user message or not.

3. Intent Classification: Intent classification uses machine learning and natural language processing to automatically equate words or expressions with a particular intent. The classifier used in this project is as follows:

- DIETClassifier : DIET (Dual Intent and Entity Transformer) is a multi-task architecture for intent classification and entity recognition. The system architecture is based on a transformer which is shared for both tasks. A sequence of entity labels is predicted through a Conditional Random Field (CRF) tagging layer on top of the transformer output sequence corresponding to the input sequence of tokens. For the intent labels the transformer output for the complete utterance and intent labels are embedded into a single semantic vector space.

4. Entity Extraction : Entity extraction is a text analysis technique that uses Natural Language Processing (NLP) to automatically pull out specific data from unstructured text, and classifies it according to predefined categories. In this project, we have used the DIETClassifier for entity extraction as well.

4. Hardware & Software Requirements

Hardware Requirements:

1. RAM: 8 GB

2. Solid State Drive

3. External Storage Capacity: 64 GB

Software Requirements:

1. Operating System: Windows 10 or above

2. Code Editor: Visual Studio Code 1.65.2

3. Programming Language: Python 3.7.6

4. Bot Framework:

- Rasa 2.2.0
- Rasa SDK 2.8.4
- Rasa X 0.34.0

5. ngrok 5.3.1

5. RESULTS

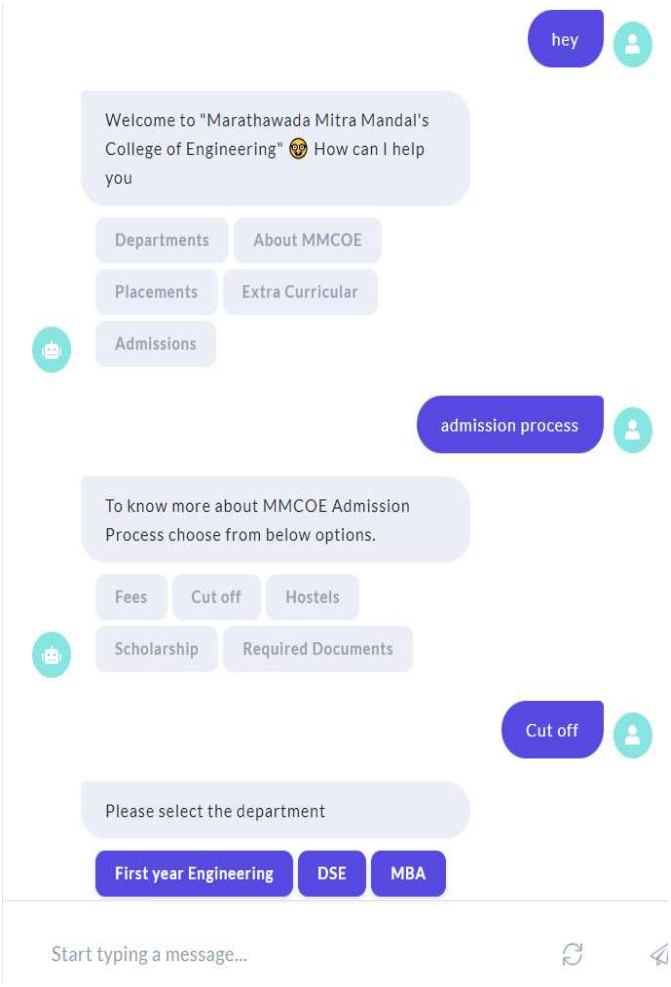

Fig -1: Web Version Output

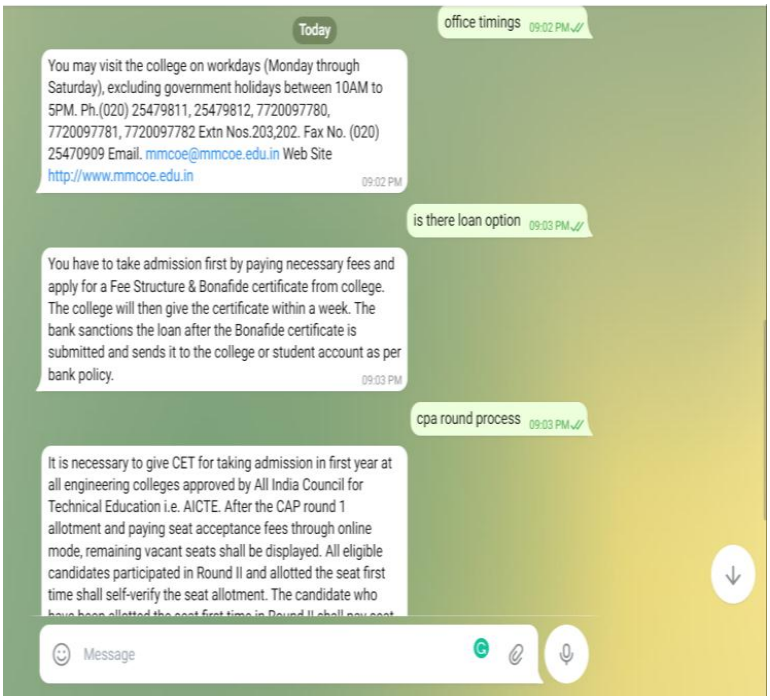

Fig -2: Telegram Version Output

6. CONCLUSION

The project proposes an AI-based College Enquiry Chatbot that could be integrated with the college website as well as the popular messaging application, Telegram to interact with visitors and answer their questions about the college. The chatbot will work with users who input their queries in both normal textual as well as custom payloads which allow users to click on their desired query instead of typing them manually.

REFERENCES

[1] Pi, S.N.M.S. and Majid, M.A., 2020, December. "Components of Smart Chatbot Academic Model for a University Website". In 2020 Emerging Technology in Computing, Communication, and Electronics (ETCCE) (pp. 1-6). IEEE.

[2] Windiatmoko, Y., Rahmadi, R. and Hidayatullah, A.F., 2021, February. "Developing Facebook Chatbot Based on Deep Learning Using RASA Framework for University Enquiries". In IOP Conference Series: Materials Science and Engineering (Vol. 1077, No. 1, p. 012060). IOP Publishing.

[3] Gbenga, O., Okedigba, T. and Oluwatobi, H., 2020. "An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot". Int. J. Comput., 38(1), pp.123-131.

[4] Kumari, S., Naikwadi, Z., Akole, A. and Darshankar, P., 2020, July. "Enhancing College Chat Bot Assistant with the Help of Richer Human-Computer Interaction and Speech Recognition". In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 427-433). IEEE.

[5] Koundinya, H., Palakurthi, A.K., Putnala, V. and Kumar, A., 2020, July. "Smart College Chatbot using ML and Python". In 2020 International Conference on System, Computation, Automation, and Networking (ICSCAN) (pp. 1-5). IEEE.

[6] Patel, N.P., Parikh, D.R., Patel, D.A. and Patel, R.R., 2019, June. "AI and web-based human-like interactive university chatbot(UNIBOT)".In 2019 3rdInternational Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 148-150). IEEE.

[7] Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N. and Gadgil, K., 2019, September. "Conversational AI: An Overview of Methodologies, Applications & Future Scope". In 2019 5th International Conference On Computing, Communication, Control An Automation (ICCUBEA) (pp. 1-7). IEEE.

[8] Ralston, K., Chen, Y., Isah, H. and Zulkernine, F., 2019, December. "A voice interactive multilingual student support system using IBM Watson". In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 1924-1929). IEEE.