**1.INTRODUCTION**

1.1 Objective-

1.2 Motivation-

The main motivation is to build a chatbot that is useful for study purposes. Stack overflow is used by tech enthusiasts for their work. So, this bot will help users to ask any question based on programming. Till now there is no such bot that can answer programming questions and which has been integrating into telegram

1.3 Background

**1. N. N. Khin and K. M. Soe, "Question Answering based University Chatbot using Sequence to Sequence Model," 2020 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), Yangon, Myanmar, 2020, pp. 55-59**.

• In the paper, the authors analysed the ways of communication through neural network chatbot by using the Sequence-to-Sequence model with Attention Mechanism based on RNN encoder-decoder model.

• This chatbot is used in the university education sector for frequently asked questions about the university and its related information.

**2. B. Setiaji and F. W. Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling," 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, 2016, pp. 72-77.**

• In this paper the machine has been embedded knowledge to identify the sentences and make a decision to answer a question.

• In this work, bigram is used for sentence similarity calculation which divides input sentences into two letters of the input sentence. The higher the score obtained the more is the similarity of reference sentences.

• The chatbot uses the knowledge that is stored in the database. The chatbot consists of core and interface that is accessing that core in relational database management systems (RDBMS).

**2.PROJECT DESCRIPTION AND GOALS**

• I will create a dialogue Chatbot, which will be able to:

• Answer programming-related questions (using StackOverflow dataset)

• Chit-Chat and simulate dialogue on all non-programming related questions

• The bot will be integrated with Telegram messenger so that we can now talk to this bot in Telegram and it’ll also be hosted on AWS.

• The bot will determine the intent and distinguish programming-related questions from general ones.

• The bot will respond to the programming question asked by tagging it with the corresponding programming language and find the most relevant Stack-Overflow Link.

• For general questions, the chatterbot will handle it.

**3.TECHNICAL SPECIFICATION**

• GOOGLE COLAB

• StackOverflow dataset.

To detect the intent of users questions we will use two datasets:

a. tagged\_posts.tsv — StackOverflow posts, tagged with one programming language (positive samples).

b. dialogues.tsv — dialogue phrases from moves subtitles (negative samples).

• CHATTERBOT, a Python library to generate automated responses to a user’s input for Chitchat-type questions for our Chatbot.

• TELEGRAM to instantiate the bot, by creating a Chatbot UI and connecting it to the telegram app backend, and run our Chatbot logic.

• AWS to host the Chatbot.

**4 DESIGN APPROACH AND DETAILS (as applicable).**

**4.1 Design Approach / Materials & Methods**

**4.2 Codes and Standards**

**4.3 Constraints, Alternatives, and Tradeoffs**

**5. SCHEDULE, TASKS, AND MILESTONES**

**(Feb 2021)**

• Understanding of the project, the objective and tool requirements, formulation of a project plan.

**(Feb-Mar 2021)**

• Creation of Intent-Classifier and Programming-Language(Tag) Classifier.

• Setting up of Telegram to make our Chatbot communicate with it.

**(Apr-May 2021)**

• To store question database embeddings to get a most similar question for the one the user has asked.

• To fit all the pieces in our SimpleDialogueManagerClass in our Telegram Bot Handler

Have created two classifiers and saved them as .pkl files.

1. Intent-Classifier: This classifier will help us predict if a question is a Stack-Overflow question or not. If it is not a Stack-overflow question, then we let Chatterbot handle it.

2. Programming-Language (Tag) Classifier: If the question is a Stack-Overflow question then this classifier will help us to predict which language(tag) a question belongs to. This is done so that we can search for those language questions in our database only.

Store Question database Embeddings

• To convert every question to a vector form and store them so we don’t calculate the embeddings for the whole dataset every time.

• Whenever a user asks a stack overflow question, use some distance similarity measure to get the most similar question.

• In essence, to have a function to get the most similar question’s post id in the dataset given, and then the given corresponding link will be shared with the user.

Telegram Bot Handler

• To fit all the pieces in our SimpleDialogueManagerClass

To integrate Chatterbot that can talk to the user and answer random chit-chat queries in our generate answer function.

To instantiate all the models, classifiers and TFIDF vectorizer objects that we’ve stored and found an index of a most similar post using the get\_similar\_question function for the generate\_answer function to respond to Stack Overflow questions

**6.PROJECT DEMONSTRATION**

Model Creation

1. Import Libraries

• Import the required libraries.

2. Read the Data

• Read the dataset files and store them as a data frame.

3. Create training data for intent classifier

• Concatenate dialogue and StackOverflow examples into one sample.

• Pre-process texts and split the data into the training set and test set in 9:1 ratio.

4. Create an Intent classifier

• Transform the train set and test set into TF-IDF features.

• Do a binary classification on TF-IDF representations of texts

• Labels will be either dialogue for general questions or StackOverflow for programming-related questions.

• Train the intent recognizer using Logistic Regression on the train set

• Check out the accuracy on the test set to check whether everything looks good.

• Dump the TF-IDF vectorizer with a pickle to use later in the running bot.

5. Create Programming Language classifier

• Prepare the data for this task and split the data into a training set and test set in an 8:2 ratio.

• Reuse the TF-IDF vectorizer that we have already created.

• Train the tag classifier using OneVsRestClassifier wrapper over LogisticRegression.

• Check out the accuracy of the test set.

• Dump the classifier to use it in the running bot.

6.

7.

Telegram Setup

• Set up a bot by talking to the BotFather in telegram and creating a name/user name for the bot.

• Will use main.py to make our Chatbot communicate with Telegram using the access token.

• main.py uses BotHandler class that implements all back-end of the bot. It has three main functions:

get\_updates - checks for new messages sent by the user.

get\_answer - computes the most relevant answer to a user's question using a SimpleDialogueManager class.

send\_message – posts the answer computed as a new message to the user.

**7. RESULT AND DISCUSSION**

We can improve on this present chatbot by increasing classifier accuracy, handling edge cases, responding faster, or adding more logic to handle more use cases.

For a chit-chat mode, we will use a pre-trained neural network engine available from Chatterbot, we can also use Seq-2-Seq models or train our own models to create such bots.

**8.SUMMARY**

**9.REFERENCES**

[1] N. N. Khin and K. M. Soe, "Question Answering based University Chatbot using Sequence to Sequence Model," 2020 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), Yangon, Myanmar, 2020, pp. 55-59.

[2] B. Setiaji and F. W. Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling," 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, 2016, pp. 72-77.

[3] L. T. Hien, L. Tran Thi Ly, C. Pham-Nguyen, T. Le Dinh, H. Tiet Gia, and L. N. Hoai Nam, "Towards Chatbot-based Interactive What- and How-Question Answering Systems: the Adobot Approach," 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh, Vietnam, 2020, pp. 1-3, DOI: 10.1109/RIVF48685.2020.9140742.

[4] Quarteroni, S. and Manandhar, S., 2007. A chatbot-based interactive question answering system. Decalog 2007, 83.

[5] Mutiwokuziva, M.T., Chanda, M.W., Kadebu, P., Mukwazvure, A. and Gotora, T.T., 2017, October. A neural-network-based chatbot. In 2017 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 212-217). IEEE.

[6] Ranoliya, B.R., Raghuwanshi, N. and Singh, S., 2017, September. Chatbot for university-related FAQs. In 2017 International Conference on Advances in Computing, Communications, and Informatics (ICACCI) (pp. 1525-1530). IEEE.

[7] Palasundram, K., Sharef, N.M., Nasharuddin, N., Kasmiran, K. and Azman, A., 2019. Sequence to sequence model performance for education chatbot. International Journal of Emerging Technologies in Learning (iJET), 14(24), pp.56-68.

[8] El Zini, J., Rizk, Y., Awad, M. and Antoun, J., 2019, July. Towards a deep learning question-answering specialized chatbot for objective structured clinical examinations. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-9). IEEE.

[9] Sreelakshmi, A.S., Abhinaya, S.B., Nair, A. and Nirmala, S.J., 2019, November. A Question Answering and Quiz Generation Chatbot for Education. In 2019 Grace Hopper Celebration India (GHCI) (pp. 1-6). IEEE.

[10] Akhtar, M., Neidhardt, J. and Werthner, H., 2019, July. The potential of chatbots: analysis of chatbot conversations. In 2019 IEEE 21st Conference on Business Informatics (CBI) (Vol. 1, pp. 397-404). IEEE.