SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

INTENT BASED CHATBOT USING RASA FRAMEWORK

INT 200 – INTERNSHIP 1 PROJECT REPORT

Submitted by

YAESHWANTH URUMAIYA - E0122028

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

(Artificial Intelligence and Machine Learning)

Sri Ramachandra Faculty of Engineering and Technology

Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai -600116

JULY, 2023

INTENT BASED CHATBOT USING RASA FRAMEWORK

INT 200 – INTERNSHIP 1 PROJECT REPORT

Submitted by

YAESHWANTH URUMAIYA - E0122028

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

(Artificial Intelligence and Machine Learning)

Sri Ramachandra Faculty of Engineering and Technology

Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai -600116

JULY, 2023

SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

BONAFIDE CERTIFICATE

Certified that this project report **"INTENT BASED CHATBOT USING RASA FRAMEWORK"** is the bonafide record of work done by **"YAESHWANTH URUMAIYA - E0122028"** who carried out the internship work under my supervision.

Signature of the Supervisor

Dr. V. Vanitha

Assistant Professor,

Department of Computer Science Engineering (Artificial Intelligence and Machine Learning),

Sri Ramachandra Faculty of Engineering and Technology,

SRIHER, Porur, Chennai-600 116.

Signature of Programme Coordinator

Dr. V. Vanitha

Assistant Professor,

Department of Computer Science Engineering (Artificial Intelligence and Machine Learning),

Sri Ramachandra Faculty of Engineering and Technology,

SRIHER, Porur, Chennai-600 116.

Evaluation Date:

SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

ACKNOWLEDGEMENT

I express my sincere gratitude to our Programme Coordinator **Prof. V Vanitha** for her support and for providing the required facilities for carrying out this study.

I wish to thank my faculty supervisor(s), **Prof. V Vanitha,** Department of Computer Science Engineering (Artificial Intelligence and Machine Learning), Sri Ramachandra faculty of Engineering and Technology for extending help and encouragement throughout the project. Without her continuous guidance and persistent help, this project would not have been a success for me.

I am grateful to all the members of Sri Ramachandra Faculty of Engineering and Technology, my beloved parents and friends for extending the support, who helped us to overcome obstacles in the study.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

This project completed during the internship is a chatbot for our college website applying the RASA framework. The objective of this project is to improve user engagement and offer effective help to website visitors. The chatbot was created to respond to a variety of user queries regarding admissions, courses, academics, and general college information. Data collecting, intent identification, dialogue management, and natural language understanding are all part of the development process. The chatbot was tested for accuracy, and general satisfaction after being trained on a dataset of frequently asked questions. Results show how the RASA-based chatbot can effectively provide users with timely and pertinent information, enhancing their overall experience on the college website.

CHAPTER 1

INTRODUCTION

1.1. INTRODUCTION TO CHATBOTS

ChatBots are an interesting implementation of Artificial intelligence and Machine learning which provides better user experience and automates most of the repetitive tasks in any given industry's customer service department. These applications are designed to recognize the end user's queries and respond to them in an human-like manner, providing users with a conversational interface to interact and obtain information or services.

There are several types of ChatBots which are being used, such as rule-based systems, which follow a predefined set of rules and follow a certain conversational path to produce responses to the user.

Retrieval-Based Chatbots, on the other hand, fetch responses to the user's input from a database of responses and output the most appropriate one to the user's input.

Generative Chatbots are applications which don't rely on any rule-base nor database, but produce outputs on the fly depending on the user's input.

Regardless of the type of the Chatbot, a few steps should be taken to build those fascinating applications.

1.2. TECHNIQUES INVOLVED

To develop a Chatbot, or any machine learning/artificial intelligence model in general. A certain set of tasks has to be completed before shipping the end-product to the client/servers. Such tasks are as follows.

1.2.1 Data Collection

Data collection in artificial intelligence and machine learning is a crucial step that helps us to build models and solve real life problems and produce quality predictions.

1.2.2 Data Processing

Processing the collected data is almost always as important as collecting the data itself. It is as such because raw data is rarely ready to use for a model. Important changes have to be made to the dataset before using it on a model.

1.2.3 Model Training

After the data is collected and processed, the natural next step is to train the model with the data in question to produce satisfactory results. This is where HyperParameter Tuning comes into play, which demands favorable conditions for the model to perform at its best.

1.2.4 Evaluation

Using Loss Functions, we can find how good (or bad) our model is, at predicting the correct outputs values for the given input dataset. With this information in hand, we can make educated choices to further improve the model.

CHAPTER 2

LITERATURE REVIEW

In the process of building a basic NLU (Natural Language Understanding) model for intent prediction, several core steps are involved. Let's focus deeper into these steps:

Data Preprocessing: Data preprocessing is crucial to convert text data into a numerical form that can be effectively trained by machine learning models. Techniques such as TF-IDF (term frequency-inverse document frequency) and COUNTVectorizer can be employed to transform the raw text data into numerical feature vectors. For RASA-specific models, the DIETClassifier (Dual Intent Entity Transformer) developed by RASA is commonly used for effective intent prediction.

TF-IDF (Term Frequency-Inverse Document Frequency):TF-IDF is a widely used technique in natural language processing to convert text data into numerical form. It quantifies the importance of each word in a document relative to the rest of documents.

It consists of two components:Term Frequency (TF): Measures the frequency of a word within a document. It assigns higher weights to words that appear more frequently in a document.

Inverse Document Frequency (IDF): Measures the rarity of a word across the entire dataset. It assigns higher weights to words that are less common in the dataset.

COUNTVectorizer:COUNTVectorizer is a simple technique for converting text data into numerical form. It counts the occurrences of each word in a document and creates a sparse matrix representation of the text data. Each document is represented by a vector where each element corresponds to the count of a specific word. It does not consider the contextual meaning of the words but provides a basic representation of the text data.

DIETClassifier (Dual Intent Entity Transformer):DIETClassifier is a specific model developed by RASA for intent classification in the RASA framework. It combines intent classification and entity recognition into a single model. It takes into account the context and sequence of words to make more accurate predictions.

Model Selection: Once the data is preprocessed, the next step is to choose an appropriate model. The choice of model depends on the complexity of the problem and the available resources. For simple intent prediction tasks, models like Logistic Regression can be employed. However, for more complex tasks, neural network models, such as recurrent neural networks (RNNs) may be utilized. In certain applications, like recommendation provider chatbots, clustering algorithms like K-Means can be considered.

Model Evaluation: After training the model, it is essential to evaluate its performance. Evaluation metrics such as accuracy, precision, recall, and F1 score are commonly used to assess the model's predictive capabilities.

Model Comparison: Different models may yield varying performance depending on the specific use case. While SVM (Support Vector Machine) models have shown effectiveness in a lot of scenarios, it is important to consider the strengths and weaknesses of each model. Logistic Regression, neural network models, and clustering algorithms each have their own use cases and applicability to specific problem domains. It is worth noting that the choice of model and preprocessing techniques should align with the specific requirements and constraints of the project. Factors such as dataset size, computational resources, and real-time performance may influence the selection process. By carefully considering these factors and iteratively refining the model, one can achieve accurate and reliable intent prediction in basic NLU models.

CHAPTER 3

IMPLEMENTATION

3.1 FRAMEWORK SELECTION

The process of selecting the most suitable framework for our chatbot application had to consider various factors such as ease of use and scalability in the future. While there are several pre-trained models available, such as OpenAI's GPT series of models, opting for a framework that provides more control over the dataset and model creation was a better option to achieve the project's goals. This was particularly important for our domain-specific application within the college context. Thus, the RASA framework due to its versatility and easy to scale capabilities. RASA also offered great documentations which could be used to understand and program the applications at a faster rate.

3.2 DATA COLLECTION

Building a chatbot tailored specifically for SRET required a unique and domain-specific dataset. Since such data is not readily available on platforms like Kaggle, we undertook the task of collecting and curating our own dataset. This process involved a combination of manual effort and collaboration with others. These other sources include other SRET students, high school students, and SRET Professors. By sourcing data from multiple contributors, we were able to ensure diversity in user inputs and cover a wide range of possible questions and scenarios.

3.3 MODEL TRAINING

Once we had collected and prepared our dataset, the next step was to train the chatbot model using the RASA framework. RASA simplifies the model training process by providing terminal commands. RASA also has custom configurations for the model parameters, such as the number of training epochs, vectorization methods, to achieve optimal performance, but the default configurations worked just as good as the others. Leveraging the power of RASA's machine learning algorithms and natural language understanding components, the model is trained to recognize user intents, extract relevant information (entities), and produce the correct responses to the intents.

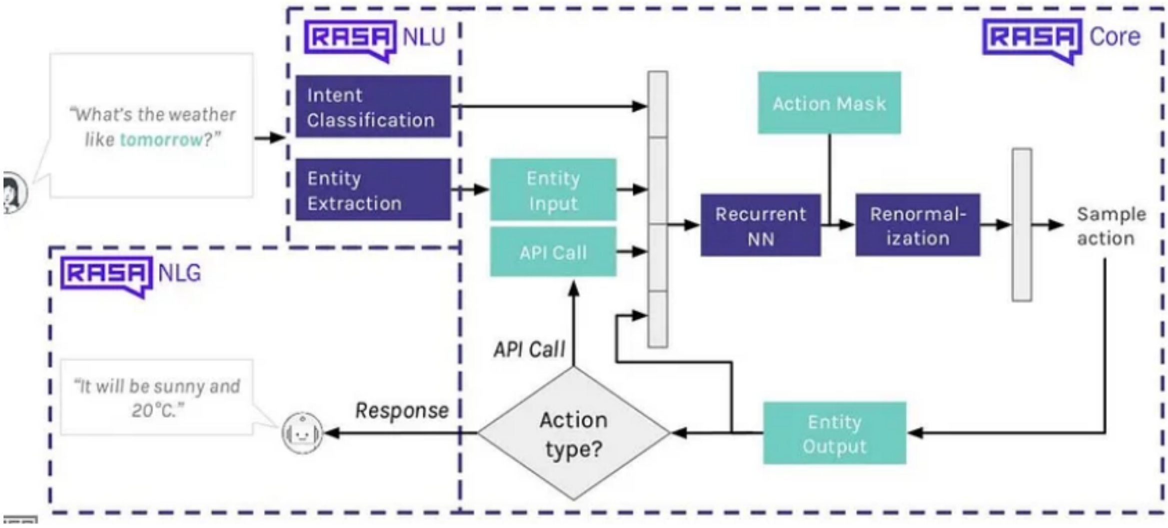

FIG 1: Process of RASA

3.4 TESTING AND EVALUATION

Testing the model is imperative. Running the command RASA SHELL will trigger a test space where you can interact with the chatbot. Testing and evaluation are essential to assess the performance and effectiveness of the chatbot. Through extensive testing in this controlled environment, we check whether the chatbot correctly recognized user intents and responded accurately to different queries and scenarios. The evaluation process involved comparing the chatbot's responses against expected outcomes and analysing metrics such as accuracy values which are presented on the command-line terminal during training.

3.5 REVALUATION OF DATA

During the testing and evaluation phase, the model had certain limitations and areas for improvement. To get rid of these issues and bugs where we have noticed during the training/testing period, we performed a revaluation of the data. This involved analysing the dataset for any potential issues such as confusing intents, misnamed response labels, or duplicate or insufficient data points. By identifying and resolving these data-related challenges, we can enhance the overall performance and accuracy of the chatbot. Expanding the dataset to include additional variations of user inputs enables the model to handle a wider range of scenarios.

Following the revaluation of the data, the next step would be retraining the chatbot model using the refined dataset. By the addition of this new data and implementation of it into the model, we improve the chatbot's understanding of user intents and its ability to identify user intents and produce correct responses.

3.6 DEPLOYMENT TO WEB

To make the chatbot accessible to users, we focused on deploying it on our college website. While RASA's default output is in the terminal, we needed to integrate the chatbot with the website's frontend to provide netter user experience. RASA's documentation provided a starting point by offering functionality to convert the terminal output to HTML output. We used this feature to generate HTML templates that would display the chatbot interface on the website. Additionally, we explored several GitHub repositories that provided upgraded versions of this functionality, specifically used together with the newer versions of RASA. Through these resources, we were able to customize the frontend appearance and user interactions, using HTML, CSS, and JavaScript to create an engaging and user-friendly chatbot interface.

By following these sequential steps, from framework selection to deployment, we successfully developed a ready-to-deploy chatbot using the RASA framework for SRET’s website. The systematic approach included dataset collection, model training, testing and evaluation, data revaluation, and deployment. The project's implementation also showcased the potential of chatbot technology in improving user experiences, streamlining communication processes, and providing valuable support within the college environment.

CHAPTER 4

RESULTS

The implementation of the chatbot using the RASA framework produced positive results. The chatbot's deployment could improve user engagement on the college website, provide efficient information on demand and reduce the amount of work for the staff members and the resources required to upkeep the website. Testing with other students, the effectiveness of the chatbot helped in improving the overall user experience. Scalability and adaptability were also kept in mind while development, allowing for the addition of an extra set of new features in the future if needed. Areas for improvement include handling complex queries and expanding the current dataset.

CHAPTER 5

CONCLUSION

In conclusion, this project successfully developed a chatbot web-application for SRET College website using RASA Framework. The chatbot's implementation process entailed the following; Framework selection, Data Collection, Model Training, Testing and Evaluation, Revaluation of Data, Deployment to Web. Through the integration of this chatbot to SRET's website, we can aim to improve user experience and provide effective support to the users.

APPENDICES

APPENDIX-1: EXTRACT OF DATASET


```
- intent: give_interested_course
  examples: |
    - yes, i am intersted in Btech [AIML](interested_course)
    - BTech [CybSec](interested_course)
    - Btech [MedSci](interested_course)
    - [AIDA](interested_course) btech
    - [aiml]{"entity": "interested_course", "value": "AIML"}
    - [aiml]{"entity": "interested_course", "value": "AIML"}
    - [medicalscience]{"entity": "interested_course", "value": "medical science"}
- intent: internship_types
  examples: |
    - what internships are offered in this college?
    - What are the different types of internships available in this college?
    - Can you tell me about the internship options?
    - What areas do the internships cover?
    - Are there specific departments for internships on this college?
    - Can you provide details about the internship programs from this college?
    - what kind of internship is offered in this college?
    - can you tell me about the internships in this college?
- intent: paid_internships
  examples: |
    - Are the internships paid?
    - Do you offer paid internships?
    - Are there any stipends for the internships in this college?
    - How much do interns get paid?
    - Are there any financial benefits for interns?
    - are the internships paid?
    - are they paid?
```

APPENDIX-2: EXTRACT OF RESPONSES

utter_ask_for_more_info_course:
- text: Do you want more specific details?

utter_info_about_arrear:
- text: Don't worry if you have an arrear or two. You can clear them by next year. As of for when to clear the said arrears. If you have an arrear from an Odd Semester; you have to clear it on an odd semester. If the arrea was in even semester; then you have to clear it on even semester. As of for the exact dates, it will announced by the prof.

utter_basic_info__course:
- text: We offer 4 Courses in BTech; AIML, AIDA, CybSec&IOT, MedicalSci. In BSc, there are 3; AIDA(HONS), Bioinfomatics, DA.

utter_ask_interested_courseForFees:
- text: Do you have any specific course in mind?

utter_ask_interested_courseForCourse:
- text: Do you have any specific course in mind?

utter_general_fees:
- text: 'BTech: 2,50,000 Rs per year. The course is for 4 Years ; BSc: 75,000 Rs per year. The course 3 years (4 years if hons)'

utter_explain_internship_general:
- text: There's two types of internships available; Internal and External. Internal is the students doing projects under the professors in the college to build their resume and External is the students doing project/work under a company. There will be internship period every year and it is wish to properly split them up into internals and externals and plan it beforehand to effective build both your resume and your work experience.

utter_explain_internship_payment:
- text: Internal internships are unpaid whereas External internship can have stipends; which depends on company that the student is interning at and the student's skillset.

utter_explain_internship_duration:
- text: There will be a quarter of an year completely dedicated to internships; Which quarter that is being allocated depends on which year the students are studying.

utter_explain_previous_companies:
- text: 'Zoro, DevRev, Smartail and Vertace are one of the many companines that our students have interned at; And in terms of Internal, We have developed various projects in industries/fields (choosen by the students). As a matter of fact, this very chatbot has been built a student during the first year internal internship.'

APPENDIX-3: EXTRACT OF CONVERSATION FLOW


```
- story: equiry about ask_about_profs
  steps:
  - intent: ask_about_profs
  - action: utter_explain_ask_about_profs
- story: interactive_story_3
  steps:
  - intent: ask_about_fees_general
  - action: utter_ask_interested_courseForFees
  - intent: give_interested_course
    entities:
    - interested_course: Bioinfomatics
  - action: Action_explain_specific_fees
  - intent: internship_types
  - action: utter_explain_internship_general
  - intent: paid_internships
  - action: utter_explain_internship_payment
  - intent: internship_assistance
  - intent: internship_duration
  - action: utter_explain_internship_duration
  - intent: previous_companies
  - action: utter_explain_previous_companies
  - intent: ask_about_arrear
  - intent: internship_competition
  - action: utter_explain_internship_competition
  - intent: nlu_fallback
  - intent: internship_assistance
  - action: utter_explain_internship_assistance

- story: interactive_story_4
  steps:
  - intent: internship_types
  - action: utter_explain_internship_general
  - intent: paid_internships
  - action: utter_explain_internship_payment
  - intent: internship_duration
  - action: utter_explain_internship_duration
  - intent: ask_about_arrear
  - intent: internship_requirements
  - action: utter_explain_internship_requirements
  - intent: internship_assistance
  - action: utter_explain_internship_assistance
```

APPENDIX-4: EXTRACT OF CUSTOM ACTIONS

```
from typing import Any, Text, Dict, List
from rasa_sdk.events import SlotSet
from rasa_sdk import Action, Tracker, FormValidationAction
from rasa_sdk.executor import CollectingDispatcher
from rasa_sdk.types import DomainDict


class Action_specific_fees(Action):
   def name(self) -> Text:
      return "Action_explain_specific_fees"
   def run(self, dispatcher: CollectingDispatcher,tracker: Tracker,domain: Dict[Text,
Any]) -> List[Dict[Text, Any]]:
      course={"aida":"2,50,000Rs for Btech and 75,000 for
Bsc","aiml":"2,50,000Rs for
Btech","cybsec":"2,50,000Rs","medsci":"2,50,000Rs","bioinfomatics":"75,000Rs",
"da":"75,000Rs"}

      inter_c=tracker.get_slot("interested_course")
      if inter_c is not None:
         inter_c=inter_c.lower()
      if inter_c in course:
         res = course[inter_c]
         dispatcher.utter_message(text=res)
      else:
         dispatcher.utter_message(text="Course fees information not found. Please
try again")
      return[]
```

REFERENCES

Journal References:

1. Johnson Kolluri, Dr Shaik Razia, Soumya Ranjan Nayak (2020), "Text Classification Using Machine Learning and Deep Learning Models", *International Conference on Artificial Intelligence in Manufacturing & Renewable Energy (ICAIMRE)*
2. Xiaoyu Luo (2021), "Efficient English text classification using selected Machine Learning Techniques", *Alexandria Engineering Journal, Volume 60, Issue 3, Pages 3401-3409, ISSN 1110-0168*

Web References:

1. RASA Learning documentation Rasa Learning Center
2. RASA General Documation Rasa: Developer Documentation Portal
3. Chatbot basics tutorial Create Chatbot using RASA | Eshaan Chauhan

WORKLOG

Day	Date	Task Done
Day 1	08/05/2023	Exploration of possible frameworks.
Day 2	09/05/2023	Introduction to RASA
Day 3	12/05/2023	Understanding RASA
Day 4	15/05/2023	Anaconda Setup
Day 5	16/05/2023	RASA Installation
Day 6	17/05/2023	RASA Installation
Day 8	19/05/2023	Research of RASA Related documentations
Day 9	22/05/2023	Research of RASA Related documentations
Day 10	23/05/2023	Learning RASA
Day 11	24/05/2023	Learning RASA
Day 12	25/05/2023	Learning RASA
Day 13	26/05/2023	Learning RASA
Day 15	02/06/2023	Drafting the project's structure.
Day 16	05/06/2023	Development of chatbot's core features.
Day 17	06/06/2023	Development of Data

Day 18	07/06/2023	Development of Data
Day 19	08/06/2023	Development of Data
Day 20	09/06/2023	Development of Data
Day 21	15/06/2023	Development of Data
Day 22	16/06/2023	Training the model
Day 23	19/06/2023	Revaluation of data
Day 24	20/06/2023	Revaluation of data
Day 25	21/06/2023	Understanding the RASA Documentation for Front-End
Day 26	22/06/2023	Usage of GitHub Repo to provide frontend functionality
Day 27	23/06/2023	Front end development (HTML)
Day 28	26/06/2023	Front end development (CSS)
Day 29	28/06/2023	Animation work
Day 30	03/07/2023	Animation work
Day 31	04/07/2023	Animation work
Day 32	05/07/2023	Report work
Day 33	06/07/2023	Report work
Day 34	07/07/2023	Presentation work