

**VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY**



Department of Computer Engineering

**Documentation for Chatbot Internship
By**

- | | |
|----------------------------|-----------------------------|
| 1. Pooja Koshti | 6.Swarangi Dali |
| 2. Dhvani Serai | 7. Omkar Mane |
| 3. Riddhi Khole | 8. Jayesh Dhanrajani |
| 4. Srishti Vazirani | 9. Sachin Mishra |
| 5. Harshita Kanal | 10. Gaurav Patil |

**Supervisor:
Dr. Anjali Yeole**

**DEPARTMENT OF COMPUTER ENGINEERING
V.E.S INSTITUTE OF TECHNOLOGY**

2019 - 2020

Problem Statement: An AI Chatbot for Vesit Admission process.

Abstract: A chatbot is an AI agent that can participate in a conversation with a user. In our project we have explored how a chatbot gives quick responses about admission related questions. Our chatbot is an artificial intelligence chatbot which gives quick responses using a ML model and also by extracting information from the college website through web scraping.

Introduction:

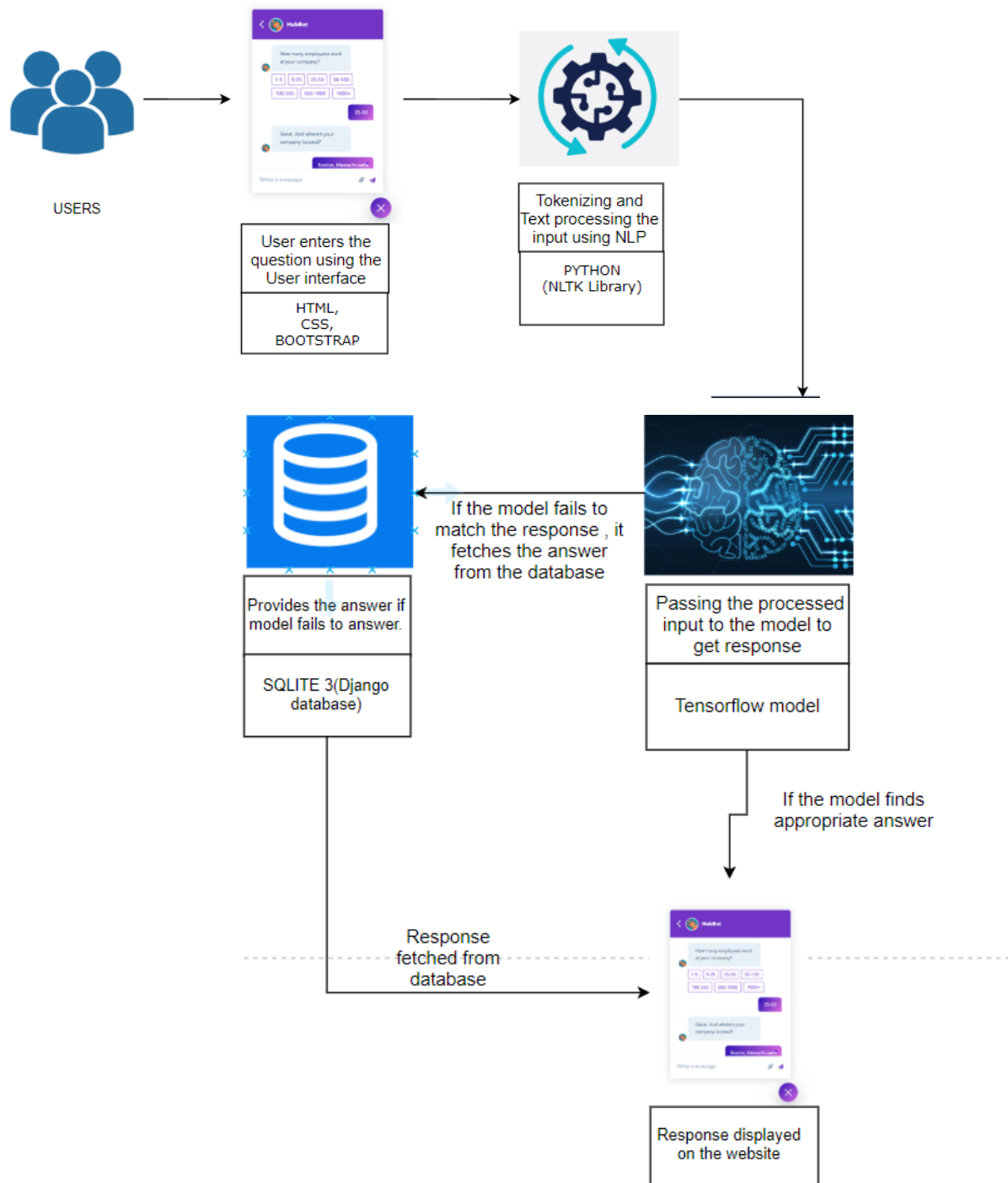
The fundamentals required for our Chatbot:

1. Natural language processing(NLP): Natural Language Processing (NLP) is concerned with how technology can meaningfully interpret and act on human language inputs. Natural Language Processing is what allows chatbots to understand your messages and respond appropriately.
2. Tensorflow and keras: Keras is an open source, high level library for developing neural network models. Basically, Keras is actually just an interface that can run on top of different Deep Learning frameworks like CNTK, Tensorflow for example.
3. Web Scraping: Web scraping, also known as web data extraction, is the process of retrieving or “scraping” data from a website. It’s extremely simple, in truth, and works by way of two parts: a web crawler and a web scraper. The web crawler is the horse, and the scraper is the chariot. The crawler leads the scraper, as if by the hand, through the internet, where it extracts the data requested.

Technologies used:

1. User Interface: Django framework with html, css, javascript.
2. Dataset: json file.
3. Machine learning: Tensorflow and keras with python.
4. Database: sqlite3 (django database).
5. Web Scraping: BeautifulSoup.

System Block Diagram:



Basic algorithm:

Step 1: Taking input from user

A button is provided to start a conversation with the chatbot. On clicking the button a window opens in which the user gives in the input . The input is fetched to the backend.

Step 2: Tokenizing the input

The input received is tokenized using Natural language processing(NLP). It gets processed and all the stop words are removed resulting in the main keywords required to pass through the Tensorflow-keras model.

Step 3: Passing through the model

The model checks for the words in the bag of words , so as to match to the appropriate response. The model used is a Sequential Tensorflow-keras model to give the responses matching to the respective tags in the dataset i.e the json file. If the input does not match any of the tags then the response is fetched from the database.

Step 4: Fetching data from Database

The database contains all the links extracted from the main college website using web scraping, so that if the model is unable to find any appropriate answer ,the database provides a respective link for the question entered by the user.

Step 5: Response is displayed to the user

The response is either given through the model or the database and is displayed on the user interface. The chatbot displays the response in a few seconds.The conversation flow is maintained and it always keeps the user in control of the conversation.

Working of the Model:

1. The dataset used to train the model is in the form of a JSON file. This file contains intents having various tags, patterns and responses to the different questions required for training the model.

```
{
  "intents": [
    {
      "tag": "greeting",
      "patterns": ["Hi", "How are you", "Is anyone there?", "Hello", "Good day", "Whats up"],
      "responses": ["Hello!", "Good to see you again!", "Hi there, how can I help?"],
      "context_set": ""
    },
    {
      "tag": "goodbye",
      "patterns": ["cya", "See you later", "Goodbye", "I am Leaving", "Have a Good day"],
      "responses": ["Sad to see you go :(", "Talk to you later", "Goodbye!"],
      "context_set": ""
    },
    {
      "tag": "fees",
      "patterns": ["What are college fees?", "What is the fees for BE?", "how much is the college fees", "fees"],
      "responses": ["1 lakh 5 thousand to 1 lakh 8 thousand. But according to different categories it differs. please it."],
      "context_set": ""
    },
    {
      "tag": "name",
      "patterns": ["what is your name", "what should I call you", "whats your name?"],
      "responses": ["You can call me Vesit bot.", "I'm vesit bot!", "I'm Tim aka Tech With vesit bot."],
      "context_set": ""
    }
  ]
}
```

2. Before training this dataset, the dataset is processed through nltk library to get the keywords and remove all the stop words, so that it becomes easy for the model to get the responses quickly. The keywords are obtained using stemming technique. We have used the LancasterStemmer for stemming.
3. The model used is a Tensorflow keras Sequential model. The model has three layers namely: the input layer, the hidden layer, the output layer. First layer 128 neurons, second layer 64 neurons and 3rd output layer contains a number of neurons equal to number of intents to predict output intent.
4. An optimizer SGD is used for improving speed and performance for training the sequential model. The model is compiled and fit with the training dataset.
5. A bag of words is created which contains the complete vocabulary extracted from the dataset. Before the model predicts the response, the input is passed through the bag of words to get the keywords so that the model predicts the response accurately.

Conclusion:

The chatbot made it easy for the user to easily get answers anytime, anywhere for their questions instead of going through the whole website. The chatbot makes the user's interaction as easy and fast as possible to ensure that the users time is not wasted and that they get what they want without any difficulty or misunderstanding from the bot.

TEAM MEMBERS:**Name - Roll no - Class.**

1. Pooja Koshti -26(D7A)
2. Dhvani Serai- 53 (D7B)
3. Riddhi Khole- 30(D7B)
4. Srishti Vazirani - 66 (D7B)
5. Harshita Kanal- 19 (D7C)
6. Swarangi Dali - 13 (D7C)
7. Omkar Mane- 32(D7A)
8. Jayesh Dhanrajani-12(D7B)
9. Sachin Mishra -23(MCA1A)
10. Gaurav Patil