Natural Language Processing

Mini Project

Title: NLP Chatbot

BE C11 - Group 1

Group Members:

Karan Adwani – 2103009
Priyanka Bajaj - 2103012
Hardik Bhagat - 2103017
Akshay Jagiasi - 2103062

Problem Statement

Project Title: NLP-Based Chatbot for Information Retrieval

This project focuses on the development of an intelligent chatbot powered by Natural Language Processing (NLP) that is specifically trained on the data available from the Thadomal Shahani Engineering College (TSEC) website. The chatbot is designed to answer user queries related to various aspects of the college, such as academic departments, faculty members, courses, subjects, and fee structures.

By leveraging NLP, the chatbot enables users to interact using natural language and quickly retrieve specific information without the need for manual navigation through the website. The project aims to enhance the user experience by providing a conversational and automated solution that simplifies the process of obtaining relevant information.

The chatbot utilizes machine learning algorithms to understand and respond to user queries accurately, drawing upon a pre-trained dataset compiled from the TSEC website. With its ability to interpret user intent, extract relevant entities, and provide precise responses, the chatbot is a practical tool for students, faculty, and visitors looking for instant answers to their questions.

Motivation

The main motivation behind developing this chatbot is to improve the user experience of those seeking information from the TSEC website. Students often need to quickly access specific details like course fees, teacher contact information, or subject details. Rather than navigating through multiple pages of the website, the chatbot can provide instant answers.

- This chatbot aims to:
- Make it easier for students and staff to get information quickly.
- Provide a more interactive experience for users.
- Automate the process of information retrieval to reduce dependency on administrative support.

- Serve as an accessible tool for potential students or parents looking for college-related information.

Problem description

Efficient Information Retrieval

Finding specific information on a website can be a daunting task, especially when it is buried within multiple sections or pages. In the context of TSEC, students or visitors might have to navigate through various pages to find information about a particular department or the faculty. This process is inefficient and time-consuming. The chatbot solves this issue by providing instant responses to user queries by accessing the relevant data directly from the TSEC website.

Understanding Natural Language Queries

A major challenge in building a chatbot is enabling it to understand human language. Users will ask questions in various forms, and the chatbot must interpret these questions accurately to provide meaningful responses. The chatbot must be able to:

- Understand different ways in which a question can be phrased.
- Recognize key terms and contexts (e.g., departments, teacher names, subjects).
- Handle incomplete or ambiguous queries and ask for clarification if necessary.

Comprehensive Coverage of Information

The chatbot must be trained on a comprehensive dataset of the TSEC website to ensure that it can answer questions from any category, such as:

- Academic departments.
- Faculty and staff information.
- Course offerings, syllabi, and subjects.
- Fees and admission processes. It is crucial that the chatbot has access to the most up-to-date information to provide accurate answers.

Handling Uncertainty and Errors

In cases where the chatbot is unable to find relevant information or misinterprets the query, it must handle the situation gracefully. It should provide suggestions for rephrasing the query or direct users to alternate sources of information.

System Requirements

Functional Requirements

The chatbot must be able to:

- Accept natural language questions from users.
- Process the input using NLP techniques to understand the intent of the query.
- Retrieve information from a pre-trained dataset that includes data from the TSEC website.
- Provide accurate and relevant responses based on the question.
- Handle a wide range of topics, such as departments, teachers, subjects, fees, and general college information.

The chatbot should also be capable of dealing with follow-up questions or clarifications to ensure an accurate response.

Non-Functional Requirements

The chatbot must be:

- **User-friendly**: The interface should be simple, allowing users to interact with the chatbot without needing technical knowledge.
- **Fast and Responsive**: It should provide quick answers to user queries, minimizing response time.

- **Scalable**: The chatbot should be able to handle multiple user interactions simultaneously without a drop in performance.
- **Accurate**: It must provide reliable and precise information based on the TSEC website's data, with minimal errors or misunderstandings.
- **Maintainable**: The chatbot's training dataset must be regularly updated to reflect any changes or additions to the TSEC website.

Design and Architecture

System Overview

The chatbot is designed using NLP techniques that enable it to understand and respond to user queries. It consists of the following core components:

1. **Frontend User Interface**: A web-based or mobile interface where users can type their questions. The interface sends the user queries to the chatbot and displays the responses. The frontend is designed for ease of use, allowing users to ask questions naturally without needing specialized syntax.
2. **NLP Engine**: The backend NLP engine processes the user's query, identifying the intent behind the question and extracting relevant entities (e.g., department names, faculty names, subjects). The NLP model is trained on a dataset built from the TSEC website content.
3. **Data Retrieval Module**: Once the intent of the query is identified, the chatbot accesses the relevant data from the dataset, which includes information such as department details, faculty profiles, subject descriptions, fees, etc.
4. **Response Generation**: Based on the retrieved data, the chatbot formulates a coherent response and delivers it back to the user. The response is designed to be concise and relevant to the user's query.

System Flow

1. **User Input**: The user submits a question, such as "What are the fees for the Computer Engineering department?"
2. **NLP Processing**: The NLP engine processes the query to identify keywords such as "fees" and "Computer Engineering department." It interprets the user's intent and recognizes the required information.

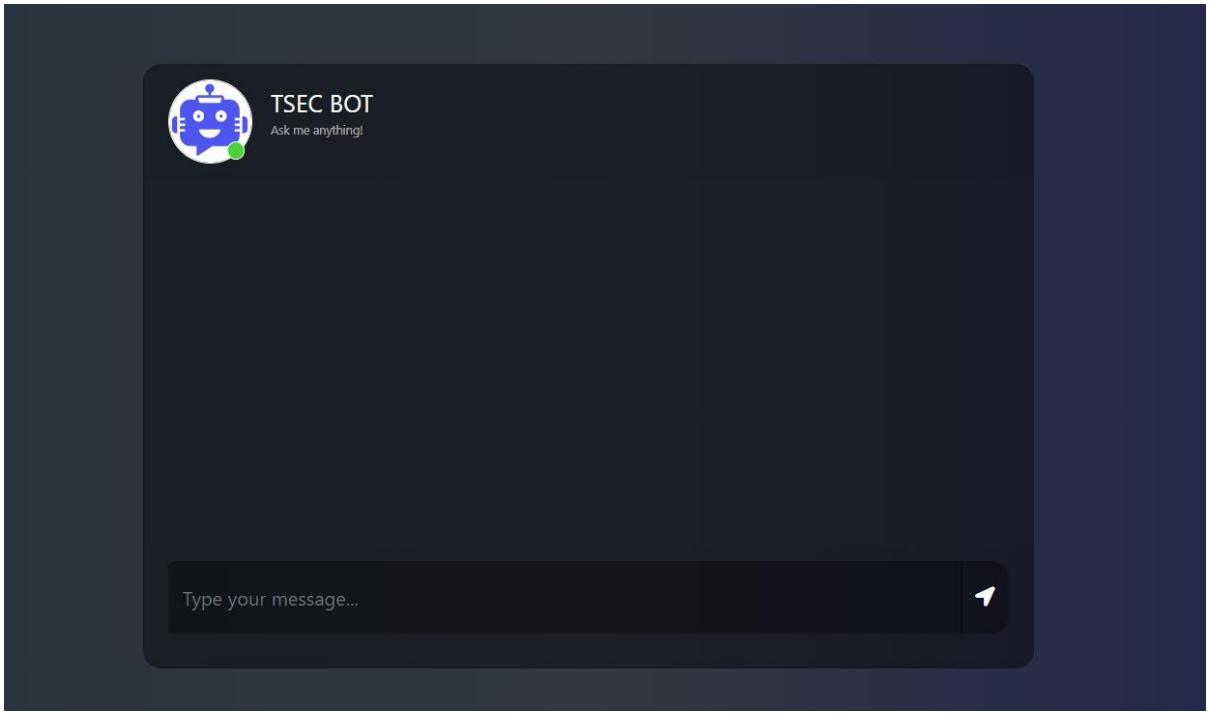
3. **Data Retrieval**: The chatbot accesses the pre-trained dataset to find the information about the fees for the specified department.
4. **Response Generation**: The system generates a response, such as “The fees for the Computer Engineering department are Rs. 1,50,000 per year.”
5. **Clarification (if needed)**: If the query is ambiguous or lacks sufficient information, the chatbot may ask for clarification, such as “Which department are you asking about?” or provide suggestions.

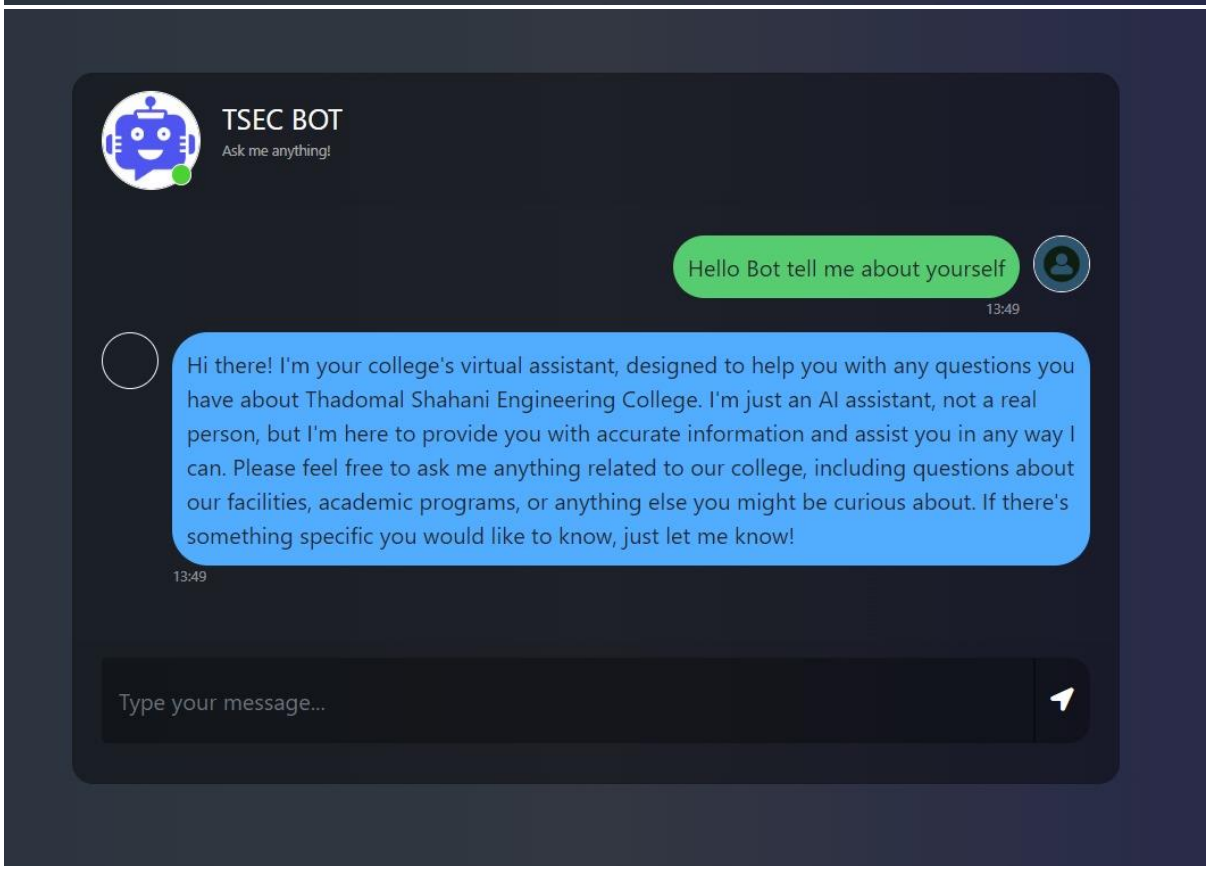
Conclusion

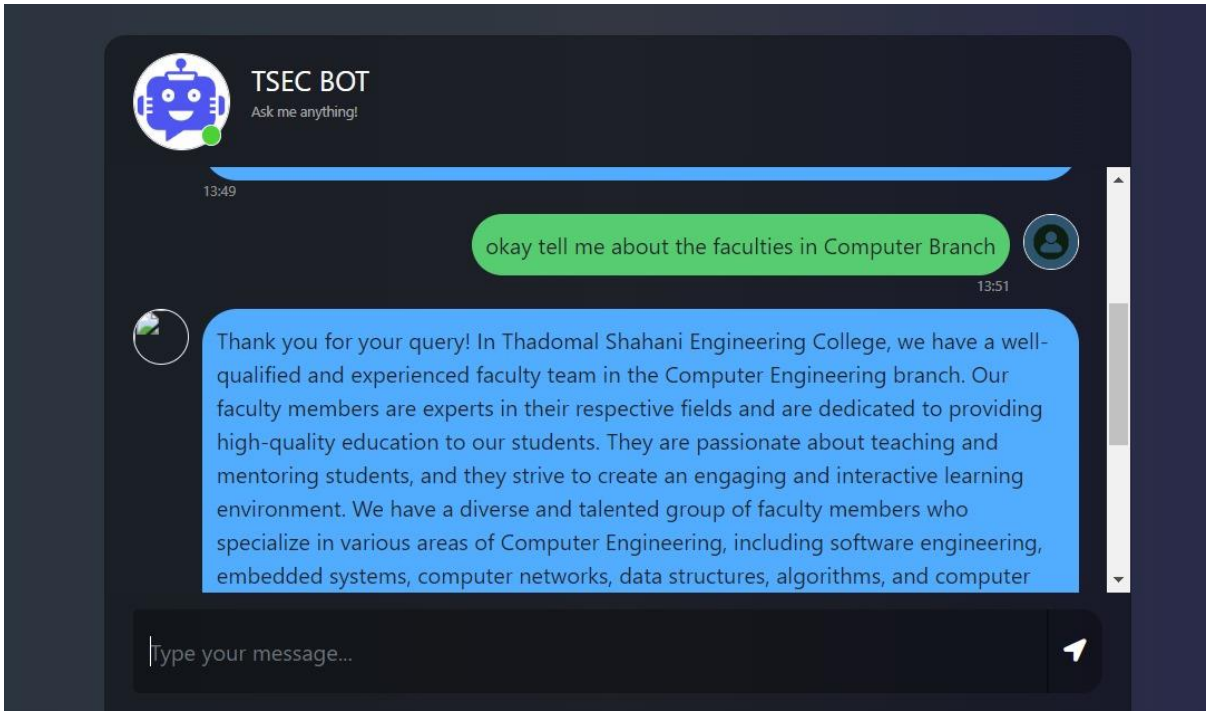
The NLP-based chatbot for the TSEC website provides an interactive, fast, and efficient way for users to retrieve information. By using advanced NLP techniques, the chatbot understands and processes user queries related to departments, faculty, subjects, and fees, among other topics.

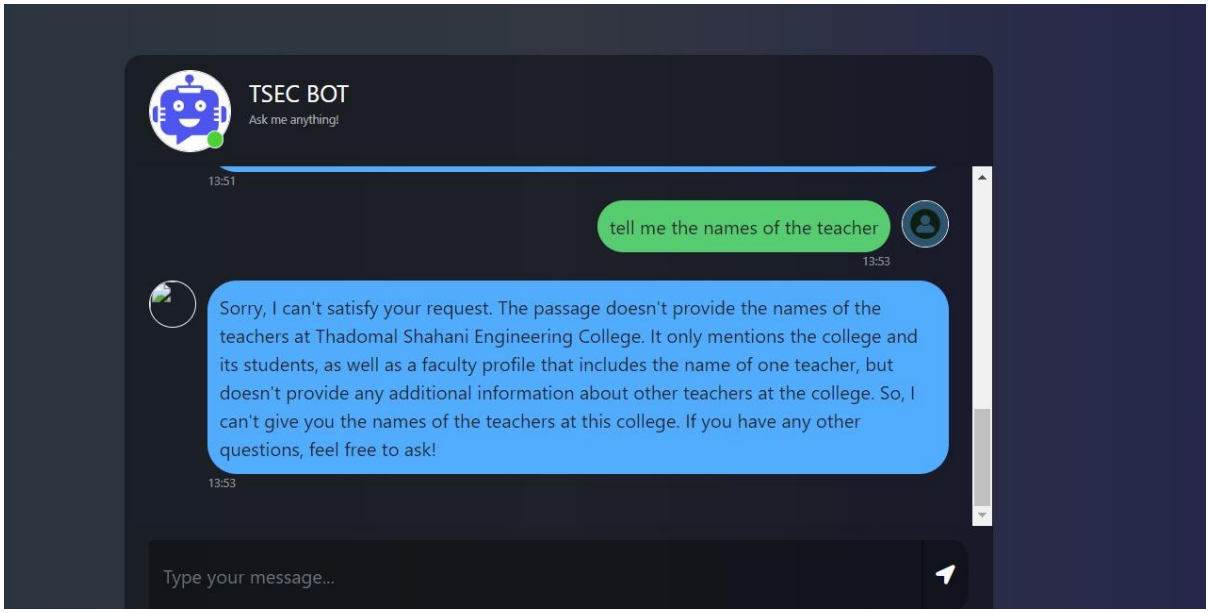
This chatbot offers a user-friendly alternative to navigating the TSEC website manually, saving time and improving the overall user experience. The project showcases how NLP can be applied to solve real-world problems by making information more accessible and easy to retrieve.

Output


TSEC BOT
Ask me anything!
Type your message...


TSEC BOT
Ask me anything!
Hello Bot tell me about yourself
13:49
Hi there! I'm your college's virtual assistant, designed to help you with any questions you have about Thadomal Shahani Engineering College. I'm just an AI assistant, not a real person, but I'm here to provide you with accurate information and assist you in any way I can. Please feel free to ask me anything related to our college, including questions about our facilities, academic programs, or anything else you might be curious about. If there's something specific you would like to know, just let me know!
13:49
Type your message...


TSEC BOT
Ask me anything!
13:49
okay tell me about the faculties in Computer Branch
13:51
Thank you for your query! In Thadomal Shahani Engineering College, we have a well-qualified and experienced faculty team in the Computer Engineering branch. Our faculty members are experts in their respective fields and are dedicated to providing high-quality education to our students. They are passionate about teaching and mentoring students, and they strive to create an engaging and interactive learning environment. We have a diverse and talented group of faculty members who specialize in various areas of Computer Engineering, including software engineering, embedded systems, computer networks, data structures, algorithms, and computer
Type your message...


TSEC BOT
Ask me anything!
13:51
tell me the names of the teacher
13:53
Sorry, I can't satisfy your request. The passage doesn't provide the names of the teachers at Thadomal Shahani Engineering College. It only mentions the college and its students, as well as a faculty profile that includes the name of one teacher, but doesn't provide any additional information about other teachers at the college. So, I can't give you the names of the teachers at this college. If you have any other questions, feel free to ask!
13:53
Type your message...