**Project Report**

**Chatbot For Department of Management BITS Pilani**

****

**Submitted By: Yash Goyal 2023H1540821P**

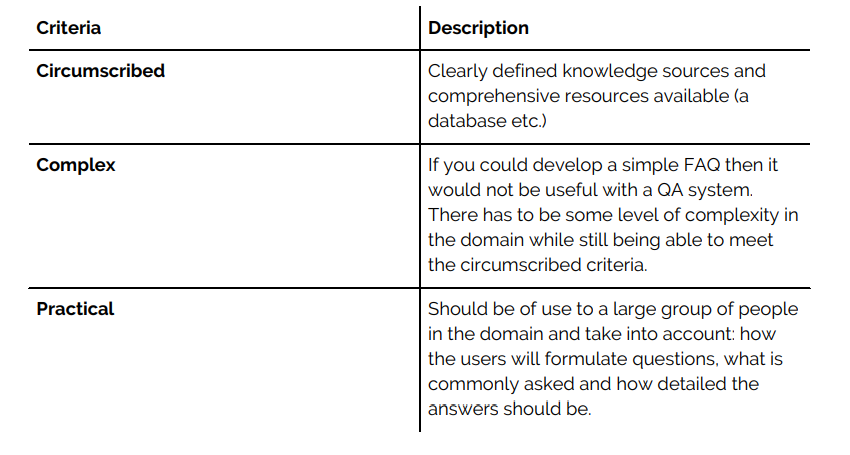
**Introduction**

Chatbots has emerged as a hot topic in the latest years, and it is used by numerous companies in various areas - help desk tools, automatic telephone answering systems, e-commerce and so on. Even though the technology has been around since the 60’s. Why are we suddenly so interested in this technology now? This can likely be explained by the recent year's advancements in messaging applications and AI technology.

... the task whereby an automated machine (such as a computer) answers arbitrary questions formulated in natural language. QA systems are especially useful in situations in which a user needs to know a very specific piece of information and does not have the time—or just does not want—to read all the available documentation related to the search topic in order to solve the problem at hand”.

Chatbots have what is called a natural language user interface and therefore communicate with users via natural language ㅡ how a human would talk on a regular basis. Therefore, they use what is called natural language processing (NLP) where the chatbot uses computational techniques to analyse text, where the goal is to produce a human-like answer based on a linguistic analysis

For a chatbot to be especially useful to a certain domain some criteria have to be met. I propose the following criteria for a domain to be successful in answering domain-specific questions:



**Overview:**

A chatbot powered by NLP can play a vital role in providing instant support, answering frequently asked questions, and guiding students through the admission process, thereby enhancing their overall experience and increasing administrative efficiency. The chatbot can assist students in selecting courses by providing details about course descriptions, prerequisites, syllabus, textbook recommendations, and available electives. The chatbot can provide information about departmental events, seminars, workshops, and guest lectures. It can also recommend relevant external resources and learning opportunities.

**Problem Statement**

‍During the admission process, students may encounter various challenges and some of them are:

**Information Overload** - The admission process typically involves navigating through a plethora of information regarding eligibility criteria, application deadlines, required documents, and fee structures.

**Clarity of eligibility requirements -** Understanding the specific eligibility criteria for programs, including academic qualifications, entrance exams, and other prerequisites.

**Communication Barriers** - Students may face difficulties in obtaining timely information from the admission office, particularly if there are communication gaps or delays in responding to queries. This lack of responsiveness can exacerbate students' anxiety and uncertainty during the admission process.

**Information Accessibility**: Students often struggle to find relevant information about courses, faculty members, events, and resources.

Addressing these challenges requires effective communication, clear guidance, and streamlined processes from the admission department.

**Collecting And Understanding data**

To understand the data collected from Admission brochure and placement report, We have first gather and organize the data into a structured format. This will involve identifying key pieces of information and remove the other information. Also, we have extracted and added information from graphs and visualisation reports. And at last we have added information based on FAQ’s and some common issues which a student faces.

We have saved all the data in a PDF in with least formatting because lot of formatting make it complex to find similarity in data and user input.

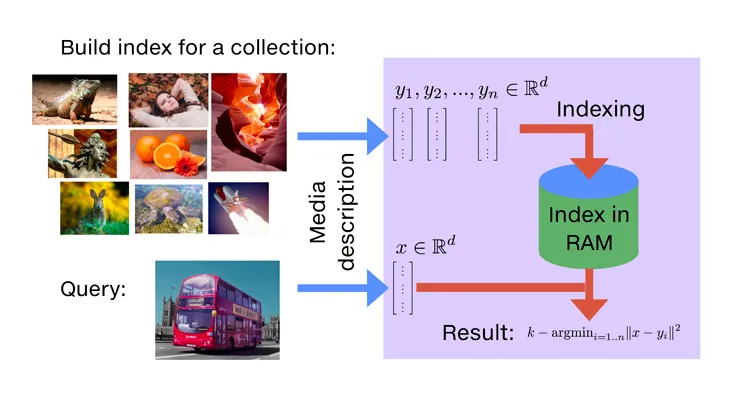
**Steps of Pre-Processing**

Pre-processing the data is an important step in preparing it for analysis. Here are some common steps that are typically involved in the data pre-processing process:

1. **Data collection :** Collecting the data from various sources such as Admission Brochure, Placement Report and Students and storing it in a structured format. Adding Answers based on previous asked questions.
2. **Text pre-processing**: This step is specific to text data and involves cleaning and transforming the text data in order to prepare it for analysis. Some common techniques used in text pre-processing include **tokenization** (breaking text into individual words or phrases).

**Faiss : Facebook AI Similarity Search**

* FAISS is used to retrieve similar document from a database based on question embedding.
* FAISS works with numerical vector representations of data. Text data, images, or other types of data need to be converted into vectors before using FAISS. Techniques like word embedding for text or image feature extraction for visuals are commonly used for this purpose.
* Give a query vector, return the list of database objects that are nearest to this vector in terms of Euclidean distance.

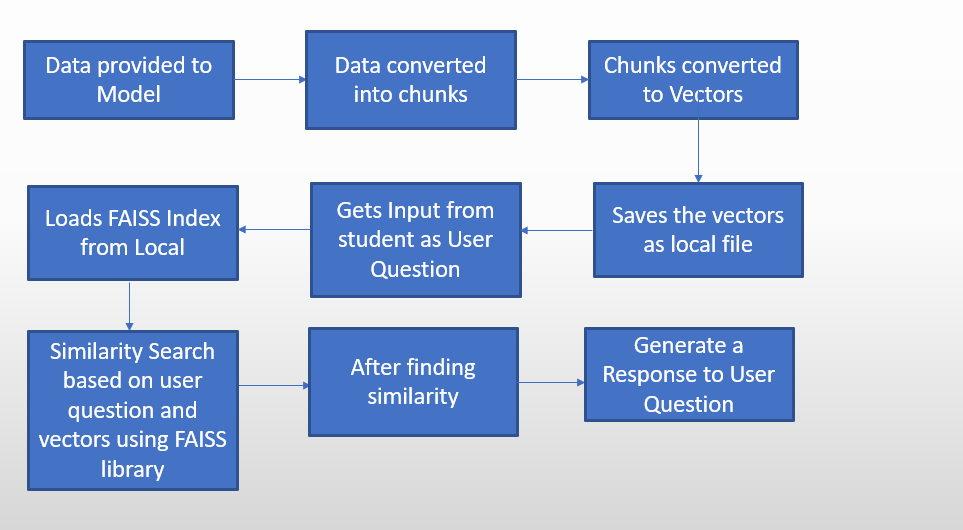


**Approach to Project**

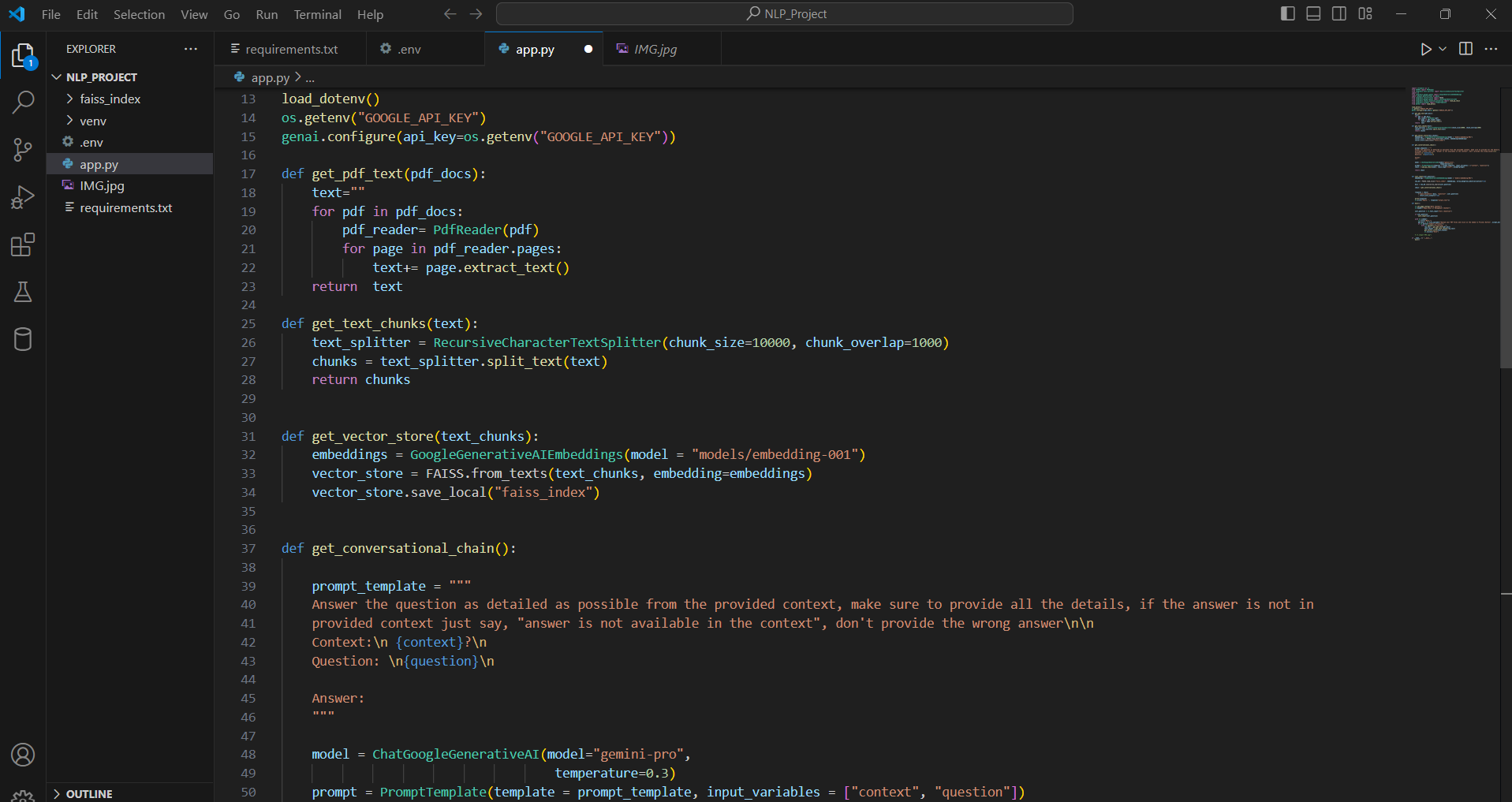
**When approaching a project involve data collection and data analysis, there are several steps that can help ensure its success. Here is a general approach that can be followed:**

1. **Problem Statement:** Start by defining the problem that the project aims to solve. This could be anything from understanding student problems and lack of availability of information. It's important to have a clear understanding of what the project aims to achieve before moving forward.
2. **Identify the data sources**: Once the problem has been defined, we have identified the data sources that will be used to gather the necessary data. This could be done by collecting information from brochure, placement reports and Previous asked queries by students.
3. **Collect and preprocess the data**: After collecting the data from brochure and previous FAQ’s. Preprocess the data as necessary, such as transforming the data into a format that can be easily analysed.
4. **Tokenization:** The bulk data is converted to small chunks of data.
5. **Vectorization:** The chunks are converted to vectors and are saved in a local file.
6. **Similarity Search:** After taking the user input the FAISS Index file is loaded and a similarity search run on the file and user input And if the similarity is found then the Response is generated.
7. **Output:** The response of the question is given to the user.

**Flow Chart**



**Code snippet:**



**Outputs:**

