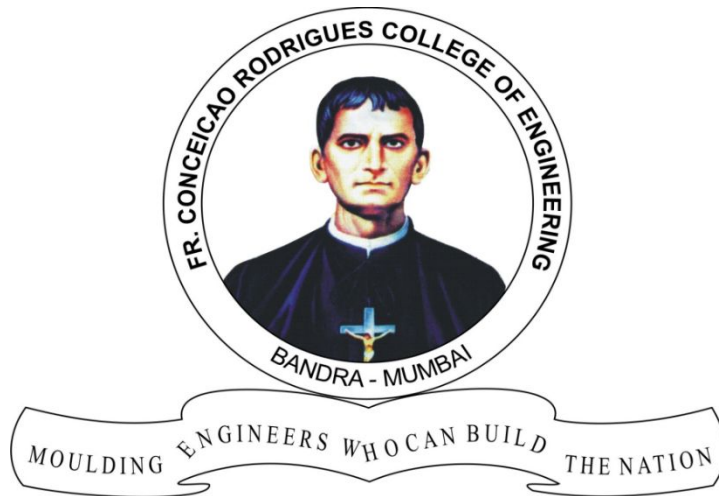


**Fr. Conceicao Rodrigues College of Engineering  
Bandra  
Computer Department  
SEM-VI**



**VIRTUAL ASSISTANT**

**Group Members:**

Princeton Baretto (8316)

Sherwin Pillai (8358)

Carol Sebastian (8320)

**Guide:**

Mrs Supriya Kamoji

---

# Table of Contents

<b>Abstract</b>	<b>1</b>
<b>1. Introduction</b>	<b>2</b>
1.1 Purpose	3
1.2 Scope	3
1.3 Definition, Acronyms, Abbreviations	4
1.4 References	4
1.5 Developer's Responsibilities	5
<b>2. General Description</b>	<b>6</b>
2.1 Product Function Overview	6
2.2 User Characteristics	7
2.3 General Characteristics	8
2.4 General Assumptions & Dependencies	8
<b>3. Specific Requirements</b>	<b>10</b>
3.1 Inputs & Outputs	10
3.2 Functional Requirements	10
3.3 External Interface Requirements	11
3.4 Performance Constraints	11
3.5 Design Constraints	12
3.5.1 Software Constraints	12
3.5.2 Hardware Constraints	12
3.5.3 Acceptance Criteria	12

## **ABSTRACT**

The project tries to develop an online platform that facilitates the existing traditional processes by introducing a virtual assistant to give support to our information technology program. We believe that the whole Question-Paper generation and assignments grading process could be a tedious job for the teachers.

Our system puts forward a solution to automatically generate questions and provide sample answers for the same from a given document. Based on these generated Q&A, it could also grade a student's handwritten answer sheet/assignment. Finding internal information from a PDF or docs should be easier. We believe that everyone should be able to use modern search technologies to find information in their documents. Our Virtual assistant allows the teachers to ask a question in natural language and get an answer without having to read the internal documents relevant to the question, making it easy for them to prepare notes for their upcoming lectures.

## **1. INTRODUCTION**

Question generation (QG) problem and Question Answering (QA), which takes a context text and a supporting phase as input and generates a question or answer corresponding to the given supporting phase, has received tremendous interest in recent years from both industrial and academic natural language processing communities. There's always a mess when teachers have to prepare Questions or any of the students or the teachers or literally any person has to understand a certain pdf or content of the text. Right from various research papers to any textbook chapter pdfs or text files, to understand or query the important questions you have in mind manually is a difficult task. Apart from this, the process of making the questions from those pdfs or text content is another difficult task since the teacher has to go through the entire file and then make or generate questions. Also, there is a lot of human thinking involved and also a lot of time is consumed in the whole process.

## **1.1 Purpose**

The system aims to develop a responsive system for Question Generation and Question Answering. The system can be used in educational institutions which would help the teachers as well as students to understand and study files easily and efficiently.

## **1.2 Scope**

AI serves to make the lives of humans easier by allowing machines to work intelligently on their behalf. However, since the technology is still in its infancy, sometimes the decisions made by artificially intelligent machines aren't fully clear and understandable as they are calculated based on a large number of variables. These variables work independently in order to achieve the best result as per their learning from the information obtained through huge data sets. One of the potential applications of AI is enquiring from a huge amount of data, as they would detect every single detail that humans think is not useful. Also through this technology we would be able to solve many doubts of students and teachers and even might be able to reduce the possibility of creating one. Further we can also connect our model to a database containing the previous data of all the data and questions asked as an additional

feature so that the user can directly browse the history. This would also help in finding the details about the previously uploaded data.

### **1.3 Definition, Acronyms, Abbreviations**

AI - Artificial Intelligence

VA - Virtual Assistant

### **1.4 References**

1. A. Agarwal, N. Sachdeva, R.K Yadav, V.Udandaraao, V.Mittal, "EDUQA: Educational Domain Question Answering System Using Conceptual Network Mapping", Million Sparks Foundation.
2. Darshana V Vekariya, Nivid R Limbasiya, "A Novel Approach for Semantic Similarity Measurement for High Quality Answer Selection in Question Answering using Deep Learning Methods", International conference on advanced computing.
3. Vignesh A, Monisha Devi, "Semantics-Enhanced Answer Selection in Closed-domain Question Answering System", 2016 international conference on power, information and communication.
4. Massimo Esposito, Emanuele Damiano, "Semantics-Enhanced Answer Selection in Closed-domain Question Answering System", 2016 international conference on signal-image technology.
5. Marco Pota, De Pietro, "Learning to Rank Answers to Closed-Domain Questions by using Fuzzy Logic", Natural Research Council of Italy.

## **1.5 Developer's Responsibilities**

The application deployment lifecycle typically involves the following key stages:

Initial Planning > Design > Development > Testing > Deployment > Support.

In each phase, an application developer is required to have specific knowledge and abilities to propel the phase into completion. These include activities around collaboration and project management. More specifically:

**Planning and Design** – This stage requires specific input from both project management and collaborative perspective. On the side of project management, the application developer must gather requirements that are based on the analysis of empirical data. For instance, the application developer should have access to end-user data that suggests what kind of issues the software sets out to solve. Armed with this information, they should have the analytical soft skills to deduce requirements. On the side of communication and collaboration is program design. This refers to building the architecture of the software and is generally spearheaded by the

developer though he or she must incorporate input from other key stakeholders on the team.

**Development and Testing** – This is where the fun begins! Based on the design requirements gathered in the previous step, the developers will begin the iterative steps of coding, testing and revising. As a leader in communication, the software developer must also work collaboratively with a panel of end-users to ensure quality assurance. Development and testing are the longest phases in the life cycle.

**Deployment and Support** – When the project is complete and tested, it will be deployed. At this point, the software developer has completed the major phases of the life cycle, and the project management focus shifts to customer support. At the same time, the developer must collaborate with other senior members of the IT department to assure asset management line items are being completed.

As mentioned above, an application developer's job doesn't end at deployment. Instead, once the application is rolled out, he or she must shift into a mode where they are prepared to monitor and release updates to the edition as needed. This is necessary for the application to run smoothly, and to reduce security risks. Application developers must also understand security protocols that protect users from external threats, and stay on the cutting edge of the changing field of technology and cybersecurity. For instance, as more and more applications begin to explore the many



uses of blockchain, application developers using will need to stay abreast of the many risks associated with this disruptive technology.

## **2. GENERAL DESCRIPTION**

### **2.1 Product Function Overview**

To design and develop an efficient cloud-based system that is capable of providing support and assistance to teaching faculty.

Also to implement Handwritten Text Recognition for student's scanned assignments, Question and Answer Generating System from any given data and Closed Domain Question Answering System for answering any questions from available resources for Teachers Assistance.

The analysis of the problem statement depends on various factors. For the proper analysis, the model needs to be trained with efficient and accurate data. The model needs to be trained in such a manner, that it can effectively produce a quality question and predict the answer to it. Initially, the given PDF's text has to be extracted and preprocessed into tokens of sentences and fed into the Closed Domain Question Generation Model.

## 2.2 User Characteristics

Users of the system should be able to put a document/pdf as the input to our model. The document can be of any text format such as pdf/txt file, etc. The model should be able to generate questions from the document and as well as answer the questions asked by the user based on the given document.

## 2.3 General Characteristics

**Question Generation:** The model should be able to generate relevant and sensible questions from the given document. The questions generated should be valid and the answers should be present in the document too.

**Question Answering:** The model should also be able to answer the questions asked by the user by going through the document provided and should return the most relevant answer to the given question.

## 2.4 General Assumptions and Dependencies

The basic assumptions and dependencies in building the software are as below

GPU availability for training.

The answer to the Questions asked should contain in the document provided.

PDF should contain text.

No limitation over Memory and Latency.

Python 3+.

### **3. SPECIFIC REQUIREMENTS**

#### **1.1 Inputs & Outputs**

The input would be pre-trained model information along with per-existing details stored about the user based on which the program would take a decision

The Output would be the text sent based on the decision taken by the program about the answer based on the paragraph.

#### **1.2 Functional Requirements**

##### **Authentication:**

All users using the model are only allowed to use it after they have been authenticated and have logged in securely

##### **Powerful and Accurate Trained Model:**

The model should be accurate and precise in predicting and recognizing the correct answers from the given set of PDFs.

### **1.3 External Interface Requirements**

#### **User Interface:**

This application must be accessible via a cross-platform and cross-vendor website and web browser-based implementation. This web browser must support all functional requirements. This user interface further must be simple and easy to understand for users of all backgrounds. This system is also available on Whatsapp as a chatbot for ease of use.

#### **Communication Interface:**

Communications from Browser to Server and from Server to Database must be done in a secure, encrypted manner to ensure complete privacy of user data and information. Any and all data of the user must be stored in a secured, encrypted form in the database. Further, the chosen protocol used for such communication must have low latency

### **1.4 Performance Constraints**

#### **User-friendly:**

It should be users friendly. The user interface should be kept simple and uncluttered. Since different types of people interact with this process, it should be very easy for them to understand.

**Flexibility:**

Addition and removal of features in the model and its interface must be simplified and possible in the future as and when required.

**Portability:**

Cross-platform compatibility and deployment are a must for this application to have maximum deployment and effectiveness

## **1.5 Design Constraints**

### **1.5.1. Hardware Constraints**

- For training the model, a GPU is preferable.
- minimum of 2GB RAM.

### **1.5.2. Acceptance Constraints**

- The document provided should contain content in text format only, any time of scanned document would not be accepted as it cannot read the text from the document.
- The document provided cannot be images, videos.

- The language of the document should be in English.