AutoQuest: Intelligent Question Generation from Online Content

N Priyanka[1], Sanvikha S[2], Shwetha Shree N[3], Tanushree A[4], Bhagyalakshmi B[5]
[5]Assistant Professor
[1, 2, 3, 4, 5] Vemana Institute of Technology

Abstract- *AutoQuest is a web-based application which gives automatic multiple-choice questions (MCQs) fromPDF, TXT, DOCX files, and YouTube videos. The application is developed using Flask and Python. AutoQuest utilizes Google Gemini 2.0 flash model to generate MCQs. The application interface gives user the choice to upload files of format .pdf, .txt or .docx or enter YouTube URL. The user can also specify the number of questions to be generated. The application extracts text from the uploaded file or the YouTube URL and generates the MCQs. The generated MCQs contains question with four options (A-D) and the correct option is also mentioned. The application has an interesting feature that is YouTube video processing to generate MCQs. The text is extracted from PDF using pdfplumber library and python-docx is used for DOCX files. For YouTube video processing the audio from the YouTube video is extracted, speech-to-text modules are used to get the text from the audio. The extracted text is sent as prompt to Gemini API to generate MCQs. The automatic question generation reduces the manual effort and the time consumed in the traditional question generation approach. The application is user-friendly and allows the user to view and download the MCQs as both .txt file and .pdf file. Teachers, Lecturers and Trainers of Schools, Colleges or Training Institutions can utilize this application to generate MCQs for assessments and quizzes.*

Keywords- Flask, MCQ, Gemini 2.0 flash, pdfplumber, python-docx, speech-to-text modules, Gemini API.

I. INTRODUCTION

In today’s growing technology, the world needs efficient, personalized and scalable tools. Teachers, trainers and professors have faced challenges in creating MCQs manually. By using AutoQuest application users can save time and reduce manual effort. The application makes use of advanced technology like Artificial Intelligence to generate MCQs from file uploads or YouTube URL. The application is built using Flask framework and Python programming language. Gemini 2.0 flash is used to generate the MCQs from YouTube URL, TXT, DOCX and PDF file formats. The Flask application scans the uploaded files or transcribes the URL and uses NLP extracts keywords from the text to form MCQs. An interesting feature about AutoQuest is that the application can generate MCQs using YouTube URL also. Audio from the YouTube video is extracted and then the audio is converted to text using speech-to-text modules like ASR. The generated MCQs contains question along with four options. The correct answer of the question is also mentioned when the MCQs are viewed. The generated MCQs can be downloaded in both TXT and PDF format. This way it helps Teachers and trainers in generating and documenting the questions which can later be used for quizzes, assignments, assessments and more.

II. RELATED WORKS

The Automatic Question Generation(AQG) research has explored Deep Learning models such as BERT and T5 in generating appropriate questions. Google’s T5 and other projects have demonstrated the capability of generative models in framing logical and meaningful questions. In the same way, Open AI’s Chat-GPT have been widely used in innovative question generation. Despite the fact that they have limited constraints in the educational field, Systems like YouTube transcript-based summarization tools extract textual explanation from videos but cannot generate MCQs. For transcribing lectures trainers depend on speech-to-text API but when it comes to question generation they rely on traditional approach. AutoQuest stands unique by automating question generation from YouTube URL, by transcribing the video into audio and then using ASR model to convert the audio into text format, the extracted text is then used to generate MCQs.

III. DRAW BACKS OF CURRENT SYSTEM

There are very few drawbacks of the current system, which doesn’t occur very often. Since generated MCQs are from the AI, which isn’t familiar with human culture, religious beliefs and political differences as humans. AI finds it difficult in generating appropriate questions as of this in concern, Questions generated manually can be carefully designed keeping all of these in mind, which results in appropriate questions. Human intelligence in particular fields like law, engineering and medical is far more better than AI, because of experience and great knowledge, therefore manual generation of MCQs is perfect than those generated by the AI. There can be grammatical errors in the MCQs generated by AI and the options given in the MCQs can sometimes not be related to the question itself. This doesn't happen when the questions are developed manually.

IV. PROPOSED WORK

Since the drawbacks of the current system is very rare and the generated MCQs can go wrong only once in a while, we still prefer MCQ generation from AI as it reduces manual effort and the time consumed in generating large number of questions.

AutoQuest is designed to generate MCQs from file uploads or YouTube URLs. The MCQs are generated using LLM, the user gives the input that can either be a file of format PDF, TXT or DOCX or a YouTube URL. The number of questions to be generated is also specified by the user. The input collected from the user is given as a prompt to the LLM. The Large Language model used in this project is Gemini 2.0 flash. The questions generated can be viewed by the user and also can be downloaded as PDF or TXT file.

Python library called pdfplumber is used to extract text from the pdf file uploaded by the user. The library, pdfplumber is highly accurate in extracting text, tables and metadata from PDF files. It has the capability to extract text from either a page of a PDF or as a whole. It can also extract text from images and shapes present in the PDF.Another python library called python-docx is used to extract text from DOCX file. The library not only extracts text from the docx file but also can create of modify the file. Text from TXT file is extracted directly as it is a raw file.

The extracted text is given as a prompt to LLM Gemini 2.0 flash. Gemini 2.0 flash is an AI model which can handle multiple inputs, work well with large volume of data. It can break down large problems into smaller components, which makes it handle the tasks well.

The generated questions are then available for the user to view and download them as PDF and TXT file. The MCQs have a question and four options, the right option is also mentioned. If the right option is not required for assessments, the MCQs can be downloaded as TXT file and the right option can be removed.

V. SYSTEM ARCHITECTURE

Figure 1 shows the System Architecture of AutoQuest MCQ generator. The user interacts with the application interface to give file uploads or YouTube URL. Based on the input the Flask web server routes for the appropriate content processing. The text that is extracted during content processing is used to generate MCQs.

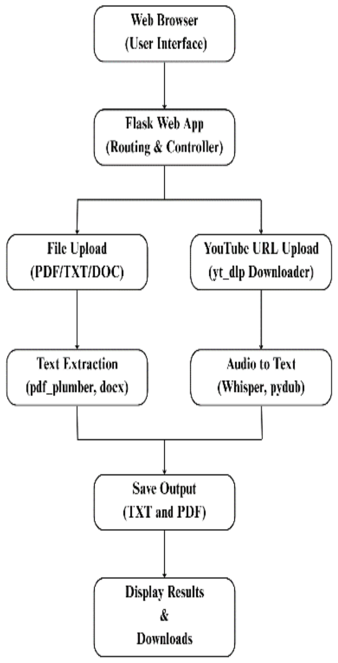


Figure 1: System Architecture of AutoQuest

VI. METHODOLOGY

1. app.py

app.py is designed to set-up the flask web app which will helps us with uploading file size limit for the document. The uploaded files can be in the form of .pdf, .txt, .docx. Then it will extract the text and uses Google Gemini for generating MCQ questions. It also accepts You Tube link as the input and download audio from video using app.2, which is external module helps it to convert and also transcripts the audio to text and generates the MCQ question from the transcribed text and we can download the question in .pdf or .txt format.

2. app2.py

app2.py is designed for downloading the audio from YouTubevideo andstores it in a temporary storage. Then the audio is converted into transcribed model. Then it will convert the entire audio file into the plain English text. After this it is imported to app.py and executes the steps.

3.index.html

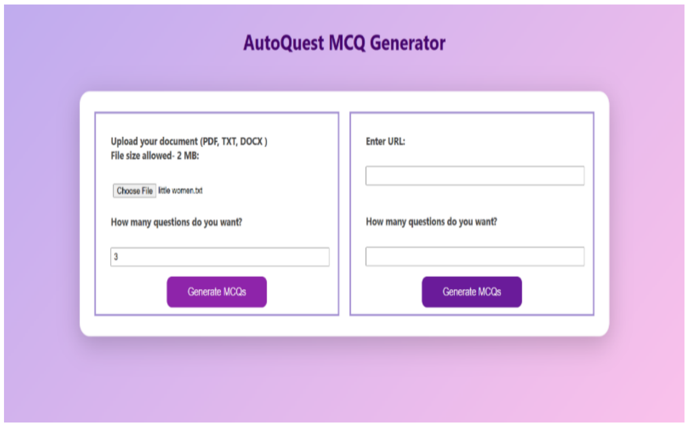
Index.html is designed to give a user-friendly interface, which is split into two sections and are aligned side by side. The first section is where the user uploads a file of maximum file 2MB. The user can also specify the number of questions to be generated. Then the second section will take the input as YouTube link and user can also specify the number of questions to be generated.

4. results.html

results.html is designed to generated the MCQ questions string in such a way that it should contain the question text, options from A to D and one correct answer. Then at last the questions can be downloaded questions in the form of .pdf or .txt. The result.html has also specified the style for the result page.

VII. RESULTS

Figure 2 shows the index page which is a user interface where the user can upload files or enter YouTube URL. The user can also specify the number of questions to be generated. Figure 3 shows the results page which is a user interface where the user can view and download the generated MCQs. The generated MCQs can be downloaded as PDF and TXT.


AutoQuest MCQ Generator

Upload your document (PDF, TXT, DOCX)
File size allowed- 2 MB:

Choose File little women.txt

How many questions do you want?

3

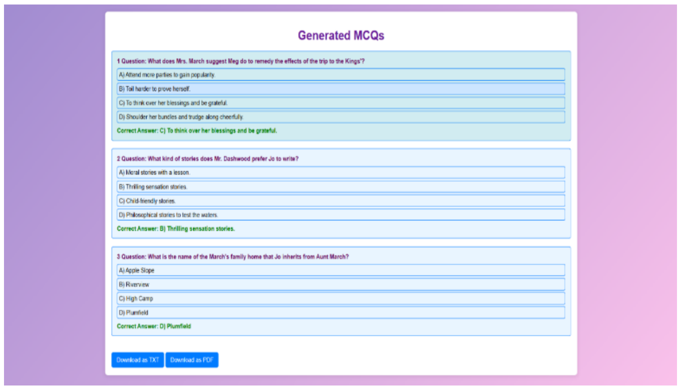
Generate MCQs

Enter URL:

How many questions do you want?

Generate MCQs

Figure 2: Index page of AutoQuest MCQ Generator


Generated MCQs

1 Question: What does Mrs. March suggest Meg do to remedy the effects of the trip to the Kings'?
A) Attend more parties to gain popularity.
B) Toil harder to prove herself.
C) To think over her blessings and be grateful.
D) Shoulder her bundles and trudge along cheerfully.
Correct Answer: C) To think over her blessings and be grateful.

2 Question: What kind of stories does Mr. Dashwood prefer Jo to write?
A) Moral stories with a lesson.
B) Thrilling sensation stories.
C) Child-friendly stories.
D) Philosophical stories to test the waters.
Correct Answer: B) Thrilling sensation stories.

3 Question: What is the name of the March's family home that Jo inherits from Aunt March?
A) Apple Slope
B) Riverview
C) High Camp
D) Plumfield
Correct Answer: D) Plumfield

Download as TXT Download as PDF

Figure 3: Results page of AutoQuest MCQ Generator

Figure 3 shows the results page which is a user interface where the user can view and download the generated MCQs. The generated MCQs can be downloaded as PDF and TXT.

VIII. CONCLUSION AND FUTURE WORK

AutoQuest is successful in generating MCQs (Multiple Choice Question) from different file uploads like PDF, TXT, DOCX and the YouTube URL. Several testes like unit testing, integration testing, functional testing, UI/UX testing, performance testing, Error handling were done to validate the functionality of AutoQuest MCQ generator. The application can generate MCQs correctly with the input 90 to 95 percentage of the time. It can handle inputs and process files up-to 2MB and YouTube videos up-to 30 minutes with the time frame 8 to 15 minutes depending on the content and the complexity of the input. Error handling gave users clear feedback when the input was invalid or of unsupported format. The application overall meets its objectives of usability and reliability.

The application can be further enhanced to support file formats like PPTs, EPUBs, OCR, images, website URLs, YouTube playlist. New features like generating MCQs in the same language as of the YouTube video. The application can also be collaborated with learning management systems for assessments and quizzes. It can further be developedto generate MCQs from lectures and podcast taking voice as input.

REFERENCES

[1] K. Bhowmick, A. Jagmohan, A. Vempaty, P. Dey, L. Hall, J. Hartman, R. Kokku, and H. Maheshwari, "Automating question generation from educational text," in *Artificial Intelligence XL: 43rd SGAI International Conference on Artificial Intelligence, AI 2023, Cambridge, UK, December 12–14, 2023, Proceedings*, Lecture Notes in Computer Science, vol. 14394, Springer, 2023, pp. 437–450, doi: 10.1007/978-3-031-47994-6_38.

[2] Zhang, "Automatic Generation of Multiple-Choice Questions," *arXiv preprint arXiv:2303.14576*, 2023. [Online]. Available: https://arxiv.org/abs/2303.14576.

[3] Zhang, "Multiple-choice question generation using large language models," *Proceedings of the 2023 ACM Conference on Educational Data Mining (EDM 2023)*,pp. 123-134, 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3631700.3665233.

[4] A. Nwafor and I. E. Onyenwe, "An automated multiple-choice question generation using natural language processing techniques," *Int. J. Nat. Lang. Comput.*, vol.

10, no. 2, pp. 1–10, Apr. 2021. [Online].Available: https://arxiv.org/abs/2103.14757.

[5] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, "A systematic review of automatic question generation for educational purposes," *International Journal of Artificial Intelligence in Education*, vol. 30, no. 1, pp. 121–204, Mar. 2020. [Online]. Available: https://doi.org/10.1007/s40593-019-00186-y.