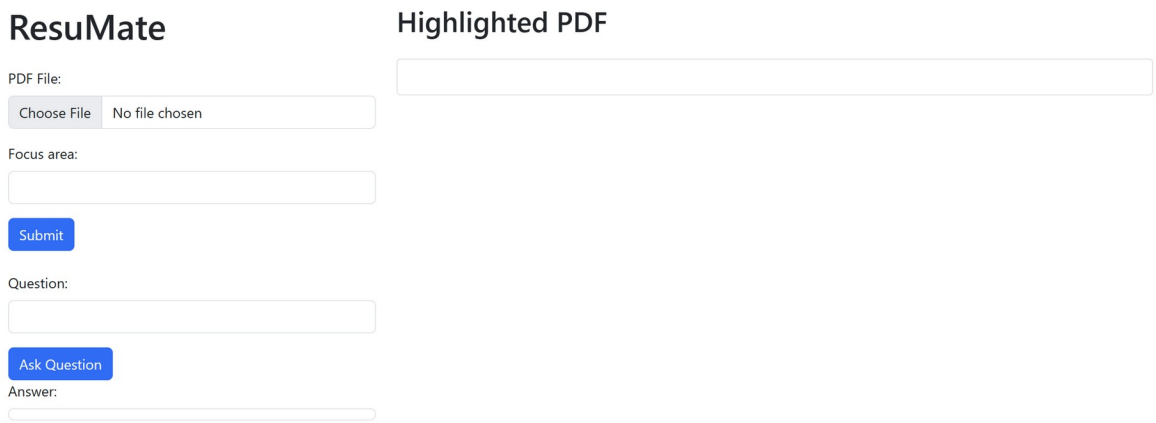
ResuMate – A Resume Highlighting System

1. Introduction

The Resume Highlighting System is a web application that allows users to upload a PDF resume and specify a focus area. The application then processes the resume, highlighting relevant texts based on the provided focus area. Additionally, users can ask questions related to the resume, and the system will provide answers based on the resume content.



The frontend of the app

2. Method

The Resume Highlighting System utilizes natural language processing (NLP) and document question-answering (QA) techniques to identify and highlight relevant information in the resume based on the user's input.

2.1 Relevant Text Highlighting

The system employs a machine learning model to identify relevant texts within the resume based on the user-provided focus area. The identified texts are then highlighted using relevance-based colored annotations in the PDF file.

2.2 Question Answering

In addition to highlighting relevant texts, the system also allows users to ask questions related to the resume. These questions are processed using a machine model which extracts relevant information from the resume content and provides concise answers.

3. Implementation Details

The Resume Highlighting System is implemented using Python and the Flask web framework. The following libraries and tools are utilized:

- Flask: The web framework used to create the application.

- PyMuPDF: A Python library for processing and manipulating PDF files. This library is used to extract text from the PDF resume, highlight relevant texts, and generate modified PDF files with highlighted annotations.
- Transformers: A library for natural language processing tasks, including question-answering. The system used pre-trained models from the Hugging Face Transformers library to identify relevant texts and answer user questions.

Model Details :

The huggingface question-answering pipeline has been used along with the deepset/roberta-base-squad2 · Hugging Face model.

The huggingface model MixedBread was implemented for token embedding and

quantified retrival. https://huggingface.co/mixedbread-ai/mxbai-embed-large-v1

Roberta-base-squad2 is the roberta-base model, fine-tuned using the SQuAD2.0 dataset. It's been trained on question-answer pairs, including unanswerable questions, for the task of Question Answering.

4. App Working

The front-end provides a user interface for uploading the PDF resume, entering the focus area and questions, and displaying the highlighted PDF files and answers

The back-end handles the following tasks:

- Receiving the uploaded PDF resume and user input (focus area and questions).
- Extracting text from the PDF resume using PyMuPDF.
- Processing the prompt and generating a set of 3-4 questions.
- Processing the resume text and identifying relevant texts based on the generated questions using the Roberta model
- Searching for and highlighting the relevant texts in the PDF resume using PyMuPDF's annotation features.
- Generating modified PDF files with highlighted annotations and saving this highlighted file.
- Processing the user question and returning the top answer based on the resume content.
- Sending the modified PDF files and answers back to the front-end for display.

5. Software Requirements

To set up the environment for running the Resume Highlighting System application, follow these steps:

- Install Python 3.6 or a later version on your system if it's not already installed.
- Install the required Python libraries (Flask, PyMuPDF, Transformers, Pillow, and fitz) using the Python package manager, pip. For example:

```
pip install flask pymupdf transformers pillow
```

- Clone or download the application's source code from the repository.
- Navigate to the application's directory and run the app.py file to start the application

```
python app.py
```

- The application should now be running, and you can access it through a web browser at the specified URL (e.g., http://localhost:5000).

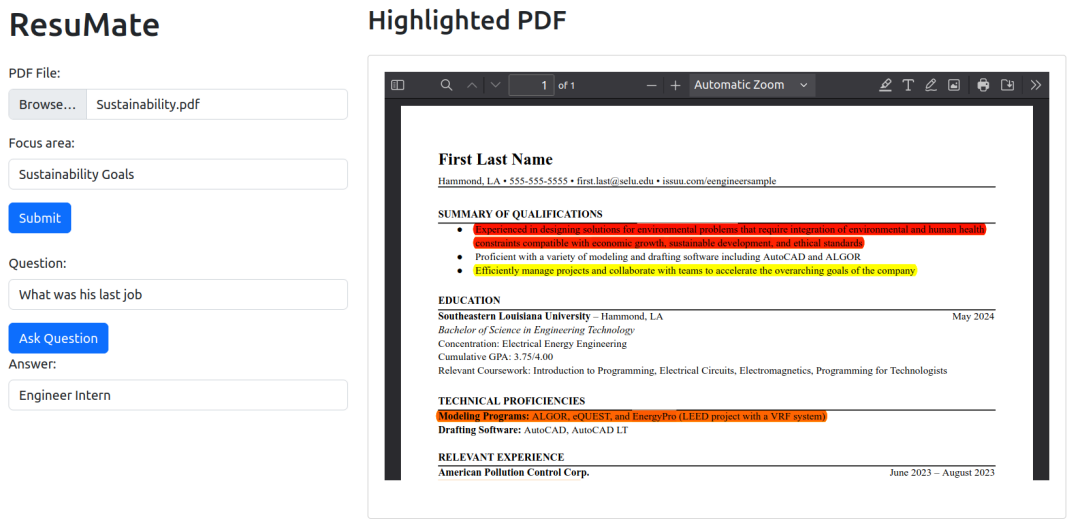
5.1 Running the app

- When the script is run for the first time, it will take 4-5 minutes to download the model from hugging face. Subsequent runs will start instantly.
- If the highlighted pdf is being downloaded but not visible on the webapp, switch to chrome/edge browser.
- The model returns better highlighted results when the input prompts are in the form of questions rather than phrases.

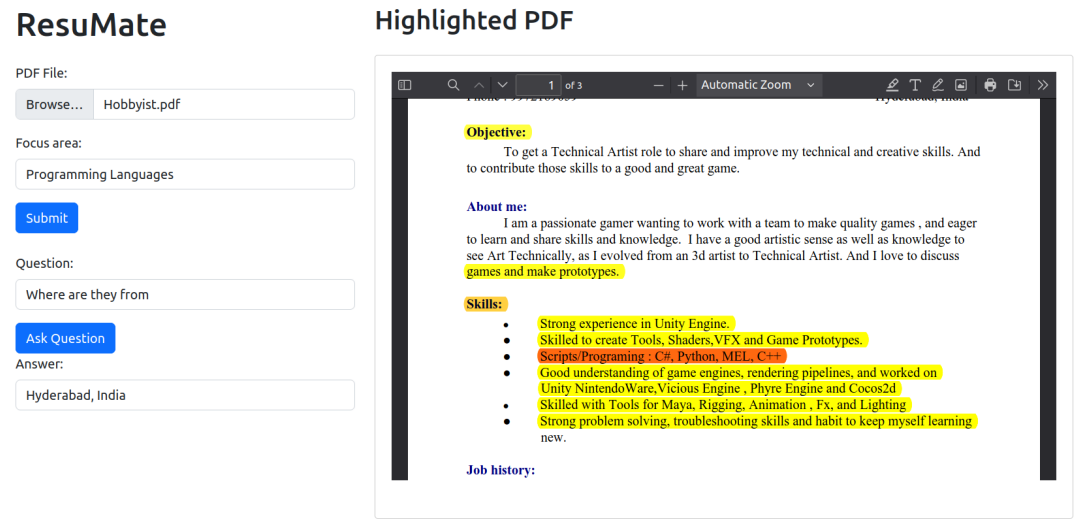
4. Performance on Test Resumes

To evaluate the performance of the Resume Highlighting System, it was tested on a set of sample resumes. The following screenshots showcase the system's output for different test cases:

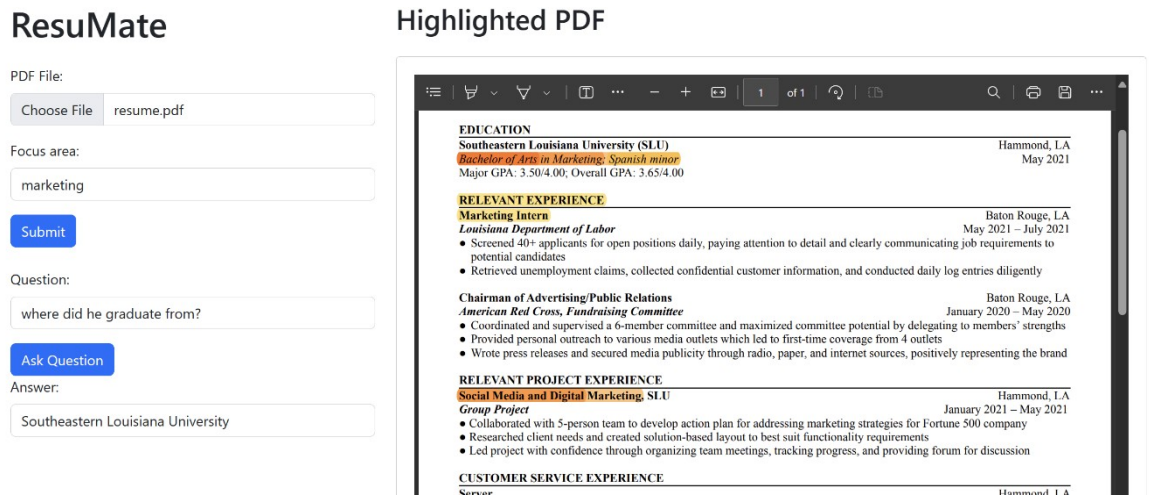
4.1 Test Case 1:



4.2 Test Case 2:

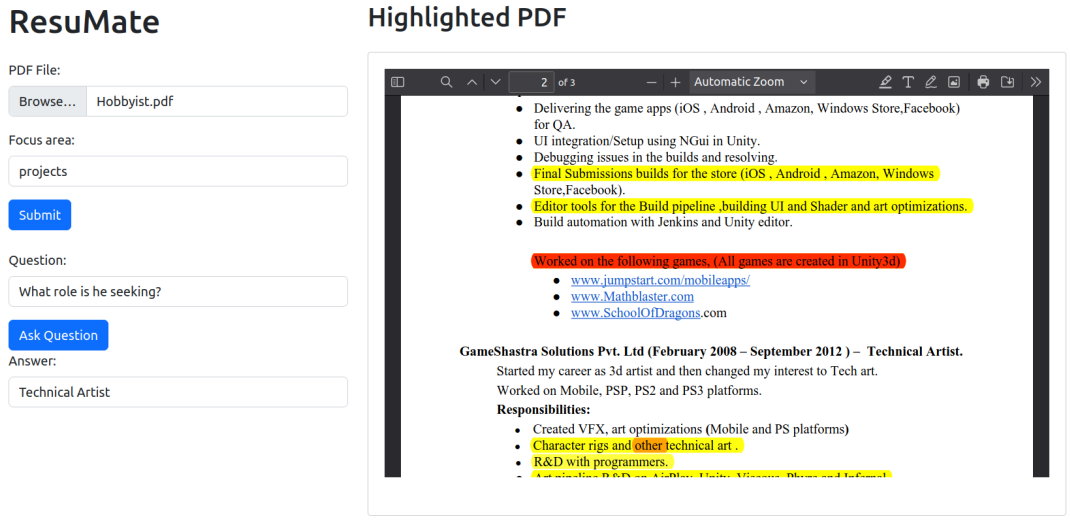


4.3 Test Case 3:



4.4 Test Case 4:

4.5 Test Case 5:



5. Conclusion

The ResuMate app is able to correctly identify and highlight the phrases (not just the keywords) in the document which are relevant to the focus area.

The app also has a complete question answering system which can taken in objective as well as subjective questions and give the answer based on the given resume.

The integration of natural language processing and question-answering techniques enables efficient information extraction and accurate response generation.