Assignment

Description:

In this task, the candidate is required to develop a web-based application in Python that allows users to upload PDF documents in Marathi language and generates a summary of the content along with extracting a list of important keywords or key phrases from the PDF. These keywords should represent the main topics discussed in the document.

Requirements:

1. The candidate should choose PDF documents in Marathi language.

2. The PDF document must have more than 15 pages.

3. Develop a frontend interface in Python where users can upload PDFs document.

4. Display the summary and extracted keywords of the PDF on the frontend interface.

Evaluation Criteria:

1. Code readability and comments.

2. Accuracy, relevancy and quality of information extraction and summary in the Marathi language.

4. Proper handling of text extraction.

5 Accuracy and relevance of extracted keywords.

6. Use of appropriate language-specific NLP techniques and Python libraries.

Submission:

The candidate should submit the following as part of their submission:

1. The Python code on GitHub.

2. A video recording: The video should include an overview of the code, running the program on the chosen PDF, and the generated results.

3. The code and any necessary instructions to run the web application locally.

Additional Note:

The frontend interface should be user-friendly and intuitive, allowing users to easily upload PDF documents and view the generated summary and keywords. The interface design and usability will be considered during evaluation.

Bonus Points (Optional):

Translation of the generated output (summary and keywords) into English. If the candidate implements a feature to translate the summary and extracted keywords from the Marathi language into English, additional points will be awarded. The accuracy and fluency of the translation will be considered during evaluation.

Deadline:

Monday, 18th March 2024, EOD