

From Text Extraction to Content Generation

Step 1: PDF Text Extraction

Process: The PDForchestration class in Java initiates the process by calling the /extract-text endpoint in the pdf_processing.py script.

Functionality: This endpoint uses PyPDF2 to read the PDF file and extract its text content.

Output: The extracted text is returned to the PDForchestration class.

Step 2: Text Summarization

Process: Once the text is extracted, PDForchestration calls the /summarize endpoint in the summarize_text.py script.

Functionality: This endpoint uses a pre-trained model from the Hugging Face Transformers library to summarize the extracted text.

Output: The summarized text is returned to the PDForchestration class.

Step 3: Bot Activation

Process: With the summarized text in hand, the PDForchestration class can now pass this information to the BotActivationService.

Functionality: The BotActivationService uses its summarizeText method to send the summarized text to the AI model endpoint specified by the summaryEndpoint.

Purpose: This step is crucial as it prepares the text to be used as a prompt for AI models, which can generate educational content, quizzes, or other AI-driven outputs.

Step 4: AI Model Processing

From Text Extraction to Content Generation

Process: The BotActivationService sends the summarized text to the appropriate AI model endpoint (e.g., content generation, quiz creation).

Functionality: The AI models process the text prompt to generate the desired output, such as educational content or quizzes.

Output: The generated content is returned to the BotActivationService, which can then be used within the application.

Summary

Integration: The PDForchestration class handles the initial stages of text extraction and summarization, while the BotActivationService finalizes the process by sending the prepared text to AI models.

Modular Design: This workflow exemplifies a modular design where each component (PDF processing, summarization, AI model interaction) is handled by specialized services, allowing for scalability and flexibility.

End-to-End Process: The entire process - from extracting text from a PDF to generating AI-driven content - demonstrates an end-to-end solution for leveraging AI capabilities in educational content creation.