Project Report

Personal Book Chapter Summarizer

1. Introduction

The Personal Book Chapter Summarizer is an AI-powered application designed to generate concise and meaningful summaries from book chapters and academic texts. The project leverages Natural Language Processing (NLP) and Retrieval-Augmented Generation (RAG) techniques to enhance summarization accuracy. A Graphical User Interface (GUI) ensures ease of access and usability for non-technical users.

2. Project Objectives

- Automate and optimize the summarization of book chapters and academic PDFs.
- Provide users with an AI-powered tool for quick and efficient content extraction.
- Enhance user experience with a simple and intuitive GUI.

3. Technologies Used

- Programming Language: Python
- Libraries & Frameworks:
  - PyPDF2 - Extract text from PDF documents.
  - transformers - Utilize NLP models for text summarization.
  - tkinter - Develop a user-friendly GUI.
  - asyncio & threading - Optimize execution for large datasets.
- Model Used: facebook/bart-large-cnn - A pre-trained summarization model.

4. System Features

- Text Extraction: Extracts text from user-selected PDF files.
- Dynamic Text Processing: Splits large text into smaller chunks to handle model input limitations.
- AI-Powered Summarization: Uses BART Large CNN model for text summarization.
- Summary Storage: Saves the generated summary as a text file for future reference.
- GUI Implementation: Provides an intuitive interface for ease of use.
- Error Handling: Ensures smooth execution even for large documents.

5. Workflow Process

1. The user selects a PDF file through the GUI.
2. The PyPDF2 library extracts text from the document.
3. The extracted text is split into smaller chunks for NLP processing.
4. The AI model processes each chunk asynchronously and generates a summary.
5. The final summary is displayed in the GUI and saved as a text file.

6. Challenges & Solutions

- Handling Large Text Data: Implemented dynamic text chunking to prevent exceeding model input limits.
- Accuracy Improvement: Utilized RAG techniques to preserve context and meaning.
- Performance Optimization: Leveraged asynchronous processing for faster execution.
- Enhancing User Experience: Designed a clean and intuitive GUI for accessibility.

7. Future Enhancements

- Multi-language support for diverse user needs.
- Compatibility with various file formats (e.g., Word, web pages).
- Voice-based input/output integration.
- Cloud-based deployment for wider accessibility.

8. Conclusion

The Personal Book Chapter Summarizer successfully automates text summarization, making it a valuable tool for students and researchers. By integrating NLP and RAG, the project ensures accurate and efficient summarization with a seamless user experience.

9. Project Details

- Developed By: Siddharth Chauhan
- Institution: Galgotias University
- Specialization: BCA (Artificial Intelligence & Machine Learning)
- Internship: Boltiot (Guidance from IIT Kanpur Professionals)