

Laboratory File

on

AGENTIC AI



School of Engineering and Technology

Department of Computer Science and Engineering

Subject code – CSCR 3215

SUBMITTED BY:

Name: Mohd Zaid

System ID: 2023307325

SUBMITTED TO:

Mr. Ayush Singh

**Sharda University
Greater Noida, Uttar Pradesh**

Lab 02: Chunking Method

Project Documentation: Multi-Level Text Splitting for Multimodal Applications

1. Project Objective

The objective of this project is to design and evaluate different **text splitting (chunking) strategies** for efficiently handling large textual and multimodal data. The focus is on improving context preservation, semantic understanding, and retrieval performance in AI-driven applications.

2. Methodology

- Input Data Collection:** Large documents such as PDFs, markdown files, code files, and other multimodal text sources are used as inputs.
- Preprocessing:** Text is cleaned and prepared before splitting to ensure consistency and efficient processing.
- Chunking Strategies:** Five levels of text splitting are implemented, including character-based, recursive, document-specific, semantic, and agent-based chunking methods.
- Tools and Frameworks:** Python, LangChain, and embedding models are used to implement and analyze the effectiveness of each strategy.

3. System Working

- The input document is analyzed and passed to the selected text splitting method.
- The text is divided into smaller chunks while maintaining logical structure and contextual meaning.
- Each chunk is stored with relevant metadata and vector embeddings.
- During inference, the most relevant chunks are retrieved to assist the language model in generating accurate responses.

4. Results and Outcomes

- Efficient processing of large documents.
- Improved contextual relevance in generated responses.
- Enhanced retrieval accuracy using semantic and agent-based chunking.
- Better understanding of the strengths and limitations of different chunking techniques.

5. Conclusion

This project highlights the importance of appropriate text splitting techniques in modern AI systems. Advanced chunking methods, especially semantic and agent-based approaches, significantly enhance system performance and are well-suited for scalable and intelligent applications.