



**DEPARTMENT OF COMPUTING
SCIENCE & ENGINEERING
LAB -2 FILE**

Course Name : Agentic_AI

Course Code: CSCR3215

Joy Kumar Paul

2023816450

SEMESTER: 6th

Section: CSH

Group: G1

FACULTY-IN CHARGE

Ayush Kumar Singh

ASSISTANT PROFESSOR

CSE DEPARTMENT

**SHARDA UNIVERSITY
GREATER NOIDA**

Documentation

Title: Fine : 5 Level of Text Splitting

This code demonstrates five different levels of text splitting used in natural language processing to prepare documents for tasks like search, summarization, and retrieval in LLM-based systems. It starts by defining a sample long text and then applies multiple splitting strategies such as character-based splitting, recursive character splitting, sentence-based splitting, token-based splitting, and semantic or structure-aware splitting using tools from text processing libraries (commonly LangChain text splitters). Each splitter breaks the same text into smaller chunks with defined chunk sizes and overlaps to preserve context between segments. The output of each method is printed and compared to show how different strategies affect chunk boundaries and context retention, helping developers choose the most suitable splitting approach for building efficient document indexing and retrieval pipelines in applications like chatbots and question-answering systems.