School of Engineering & Technology
Department of Computer Science & Technology

SHARDA
UNIVERSITY
Beyond Boundaries

Agentic AI
LAB FILE
for
B. Tech. (CSE)

BY
Rakshit Bisht(2023499171)

5 Levels of Text Splitting (Chunking)

Text splitting (also called chunking) is the process of breaking large text into smaller parts so that language models can process, store, and retrieve information efficiently. It improves performance in applications like RAG, chatbots, and document search.

Level 1: Character Splitting

- Splits text after a fixed number of characters
- Ignores meaning, sentences, or structure

Pros:
✓ Very simple
✓ Easy to implement

Cons:
❌ Breaks sentences
❌ Loses context

Use case:
Only for learning basics

Level 2: Recursive Character Text Splitting

- Splits text step-by-step using separators
- Order of separators:
    a. Paragraphs (\n\n)
    b. New lines (\n)
    c. Spaces
    d. Characters

Pros:
✓ Keeps sentences together
✓ Better context preservation

Cons:
❌ Still size-based

Use case:
Articles, essays, blogs, notes

Level 3: Document-Specific Splitting

Splitting depends on document type.

Markdown

- *Splits by headings (#, ##)*
- *Keeps sections intact*

Python Code

- *Splits by class, def*
- *Preserves functions and logic*

JavaScript

- *Splits by function, const, let, loops*

Pros:
✓ *Structure-aware*
✓ *Best for code & formatted docs*

Use case:
PDFs, Markdown files, source code

Level 4: Semantic Splitting

- *Uses embeddings (meaning) instead of characters*
- *Groups text by topic similarity*

Pros:

✓ *Best context quality*

✓ *Ideal for Q&A and RAG*

Cons:

❌ *Computationally expensive*

❌ *More complex*

Use case:

AI search, knowledge bases, RAG systems

Level 5: Agentic Splitting

- *AI agent decides how and where to split*
- *Dynamic and adaptive*

Pros:

✓ *Most intelligent approach*

Cons:

❌ *Experimental*

❌ *Slow and costly*

Use case:

Advanced AI agents & research systems