

Modern data-driven applications rely heavily on text processing pipelines.

Text extraction is the first step in converting documents into structured data.

Once extracted, the text is divided into smaller logical chunks.