

LangChain PDF Splitting Test Document

Section 1: Introduction

LangChain is a framework designed to simplify the creation of applications using large language models. This document is intended for testing text splitters such as RecursiveCharacterTextSplitter and TokenTextSplitter.

Section 2: Long Paragraph

Python virtual environments help isolate dependencies. When working with embeddings, chunking text correctly is essential to maintain semantic meaning across splits. This paragraph is intentionally long to test chunk boundaries and overlaps. Chunk overlap helps preserve context across splits and improves retrieval accuracy in vector databases.

Section 3: Bullet Points

- LangChain
- Text Splitters
- Embeddings
- Vector Stores
- Retrieval

Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests. Section 4: Repeated Content

This section is repeated multiple times to generate sufficient length for splitting tests.

Section 5: Second Page Content

This page exists to test multi-page PDF loading and splitting behavior. PDF loaders should correctly extract text from multiple pages without losing order.