

mdya PDF 取り込み日本語スモーク用フィクスチャ

tests/smoke_ingest.rs から呼ばれます。

pdf-extract が CMap / ToUnicode を

正しく解釈できるかを確認するためのドキュメントです。

短い段落を複数置いて chunker の挙動を観察します。

検索キーワード: フィクスチャ / スモーク