Test Document

Introduction

This is a test document created to demonstrate the PDF parsing capabilities of the vlm_doc_test library. It contains various elements including headings, paragraphs, and structured content.

Methods

The parsing process uses PyMuPDF to extract text with coordinate information, detect structural elements like headings based on font size, and identify visual elements such as figures and tables.

Results

The parser successfully extracts content elements with bounding box information, enabling coordinate-aware analysis and comparison with VLM-based extraction methods.