

Complex Document Structure Test

A test document with multiple levels of headings

1. Introduction

This document demonstrates a complex structure with multiple heading levels. It is designed to test the heading detection algorithm.

1.1 Background

This is a background section with some context for the document.

1.2 Purpose

The purpose of this document is to test heading detection.

1.2.1 Primary Goals

- Test H1 detection
- Test H2 detection
- Test H3 detection

1.2.2 Secondary Goals

Ensure proper hierarchical structure is maintained throughout the document.

2. Methodology

This section describes the methodology used in the document.

2.1 Approach

The approach involves creating a PDF with various heading formats.

2.2 Tools Used

The following tools were used in this document:

- PyMuPDF
- ReportLab

Implementation Details

This section has no numbering to test detection by font size only.

Technical Specifications

This subsection also has no numbering to test detection.

RESULTS AND DISCUSSION

This section uses all caps to test capitalization detection.

Key Findings

This subsection lists the key findings from the analysis.

• Finding 1:

Bullet points with bold and colon may be detected as H3.

• Finding 2:

Multiple similar patterns help establish consistent detection.

Conclusion

This document demonstrates various heading styles and formats to test detection algorithms.

REFERENCES:

1. Smith, J. (2022). PDF Structure Analysis. Journal of Document Processing.
2. Lee, K. (2023). Heading Detection in PDFs. Digital Library Research.