

PDF Outline Extractor

1. Introduction

The PDF Outline Extractor is a tool designed to analyze and extract the table of contents or outline structure embedded within PDF files. This is useful for creating summaries, navigation aids, or content indexing.

2. Features

- Extract hierarchical outline structures
- Display chapter titles and page numbers
- Support for encrypted PDFs
- Export outline data to JSON or plain text

3. Use Cases

Researchers, students, and professionals can use this tool to quickly navigate complex PDF documents or automate content indexing. It is also useful for digital libraries and document management systems.

4. Technologies Used

- Python
- PyPDF2 / pdfminer.six / fitz (PyMuPDF)
- FPDF for report generation
- JSON for structured output

5. Conclusion

PDF Outline Extractor is a powerful yet simple utility that streamlines the process of retrieving document structure. Its versatility and ease of integration make it suitable for a variety of document-processing workflows.