CHONKER & SNYFTER Test Document

Generated on: 2025-07-15 21:24:11

This is a test document created to debug the CHONKER & SNYFTER application. It contains various types of content including headings, paragraphs, lists, and tables to test the document processing capabilities.

1. Features of CHONKER

- PDF Processing - Extracts content from PDF files
- De-scuzzifying - Cleans up messy PDFs
- Chunk Extraction - Breaks documents into manageable pieces
- Batch Processing - Handle multiple files at once

2. Features of SNYFTER

- Database Storage - Archives all processed documents
- Full-Text Search - Find content quickly
- Export Options - JSON, CSV, Markdown formats
- Metadata Tracking - Keep track of processing details

3. Processing Statistics

Document Type	Processing Time	Success Rate
Simple PDF	0.5 seconds	100%
Complex PDF	2.3 seconds	95%
Scanned PDF	5.1 seconds	85%
Annotated PDF	1.2 seconds	98%

Technical Implementation

Architecture Overview

The system uses a modular architecture with the following components:

1. PDF Processing Engine (CHONKER)
    - Uses PyMuPDF for PDF manipulation
    - Docling for content extraction
    - Multi-threaded processing for performance

2. Database Layer (SNYFTER)
    - SQLite with FTS5 for full-text search
    - Structured schema for documents and chunks
    - Efficient indexing for fast retrieval

3. User Interface
    - PyQt6 for modern, responsive UI
    - Dark theme with animations
    - Real-time progress feedback

Code Example

```
# Process a PDF with CHONKER
worker = EnhancedChonkerWorker(pdf_path)
worker.finished.connect(on_processing_complete)
worker.start()

# Search with SNYFTER
results = db.search_documents('machine learning')
for doc in results:
    print(f'Found: {doc["filename"]}')
```