

Test Document for RAG System

This is a test document to demonstrate PDF parsing capabilities.

The RAG system should be able to extract:

- Text content from multiple pages
- Document metadata (title, author, creation date)
- Page structure and numbering

Customer Support Scenario:

A user reported login issues with the system. The problem occurred when trying to access the dashboard after a recent security update. The user received an error message: 'Authentication failed'.

Resolution steps:

1. Clear browser cache and cookies
2. Reset password using the forgot password link
3. Try logging in with a different browser
4. Contact IT support if the issue persists

Technical Documentation

System Requirements:

- Python 3.9 or higher
- ChromaDB for vector storage
- Ollama for local embeddings

Configuration:

The system uses JSON configuration files to define:

- Parser settings (CSV, PDF)
- Embedder configuration (model, batch size)
- Vector store settings (collection name, persistence)

API Endpoints:

- /ingest - Process and store documents
- /search - Query the knowledge base
- /info - Get system status

This document contains multiple paragraphs and structured information that should be properly extracted by the PDF parser.