Enterprise Document Management System

This document outlines the key features and requirements for an enterprise document management system that supports automated ingestion and indexing.

# Key Features

* Automated document ingestion from PDF and DOCX formats
* Intelligent text extraction and cleaning algorithms
* Semantic search and retrieval capabilities
* Source citation with original document links
* RESTful API for integration with existing systems

# Implementation Requirements

The system must handle multipart file uploads, process documents using appropriate libraries, extract clean text content, and automatically index the processed content into a searchable knowledge base. When users query the system, responses should include direct links to the original attachments for verification and reference.

## Technical Implementation

The document processing pipeline uses industry-standard libraries for text extraction, implements chunking strategies for optimal indexing, and maintains strong associations between processed content and source documents.