Dana AI Knowledge Base System

Technical Documentation

# Introduction

The Dana AI Platform provides a comprehensive knowledge base system that allows organizations to store, organize, and retrieve information effectively. This document outlines the key components and functionality of the system.

# System Components

The knowledge base system consists of the following core components:

* File Storage and Management: Handles uploading, versioning, and organization of files.
* Content Extraction: Parses various file formats to extract searchable text and metadata.
* Search Engine: Provides full-text and semantic search capabilities across all content.
* API Interface: Offers RESTful endpoints for programmatic access to knowledge content.
* Integration System: Connects with external platforms and services for data exchange.

# Supported File Formats

|  |  |  |
| --- | --- | --- |
| Format | File Extension | Features |
| PDF Documents | .pdf | Full text extraction, metadata, page-by-page content |
| Word Documents | .docx | Text extraction, metadata, paragraph-level content |
| Plain Text | .txt | Full text import, line-by-line processing |
| Rich Text Format | .rtf | Formatted text extraction |

# Integration Capabilities

The knowledge base system integrates with various components of the Dana AI Platform, including:

* Conversation System: Provides relevant information for automated responses.
* User Dashboard: Allows users to browse and search knowledge content.
* Analytics Engine: Tracks knowledge base usage and effectiveness.
* External APIs: Connects with third-party services and data sources.

# Implementation Details

The knowledge base system is implemented using the following technologies:

|  |  |
| --- | --- |
| Component | Technology |
| Database | Supabase (PostgreSQL) |
| File Storage | Supabase Storage |
| Search Engine | PostgreSQL Full-Text Search |
| API Layer | Flask RESTful API |
| File Parsing | Custom Python Utilities (PyPDF2, python-docx) |

# Security Considerations

The knowledge base system implements several security measures to protect sensitive information:

* User Authentication: All access requires valid authentication credentials.
* Row-Level Security: Content is isolated at the database level by user ID.
* Access Control: Granular permissions control who can view, edit, or delete content.
* Content Validation: Uploaded files are scanned and validated before processing.
* Audit Logging: All operations are logged for security monitoring and compliance.

# Conclusion

The Dana AI Knowledge Base System provides a robust foundation for storing, organizing, and retrieving information across the platform. It enables organizations to leverage their existing documentation and knowledge assets to enhance customer interactions, streamline operations, and improve decision-making processes.