AI-Powered PDF Knowledge Base

Semantic Search and LLM Question Answering

Team Name: *Ignite Team*

College Name: *SRI SHAKTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY*

Department: *Information Technology*

Guide Name: *Mrs. M. Deepa AsP/IT*

The Challenge: Navigating Information Overload

Inefficient PDF Navigation

Traditional methods struggle with large, complex PDF documents, leading to significant time loss in information retrieval.

Keyword Search Limitations

Keyword-based searches often fail to capture the semantic meaning, missing relevant information and providing irrelevant results.

Need for Intelligent Understanding

There's a critical demand for AI-driven solutions that can truly comprehend document content and provide precise answers.

PROJECT GOALS

Defining Our Vision: Key Objectives

Semantic Search Development

To build a robust semantic search system capable of understanding context and meaning, beyond just keywords.

AI-Powered Q&A

To enable natural language question answering using Large Language Models (LLMs) for direct, concise responses.

Privacy and Offline Support

To ensure user data privacy and offer seamless functionality even without an active internet connection.

Our Solution: The AI PDF Assistant

Intelligent PDF Interaction

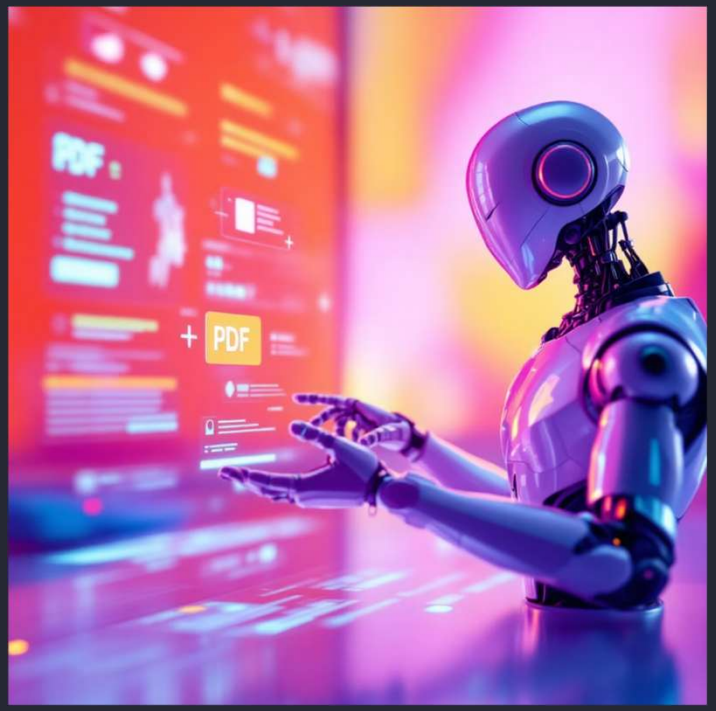
An AI-powered assistant designed to transform how users interact with PDF documents.

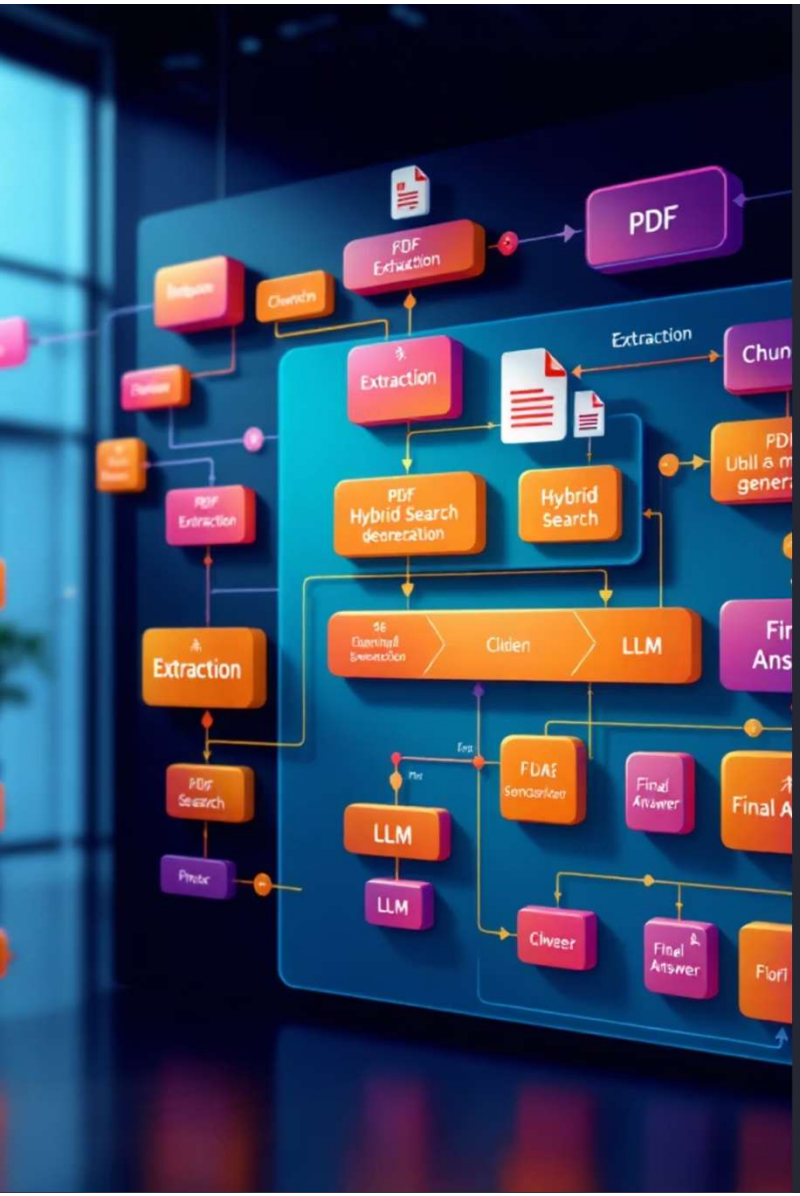
Hybrid Search Engine

Leveraging both traditional keyword and advanced semantic search techniques for comprehensive results.

Flexible Deployment

Support for both local, on-device processing and cloud-based scalability to suit diverse needs.





TECHNICAL FLOW

System Architecture: A Unified Approach

PDF Input | Extraction | Chunking | Hybrid Search | LLM

This architecture integrates various components, from document preprocessing to AI-driven response generation, ensuring a seamless information flow.

Core Components: Building Blocks of Intelligence

PDF Processor

Handles document parsing, text extraction, and initial data preparation.

Vector Store

Stores vectorized document chunks for efficient semantic similarity searches.

Hybrid Search

Combines keyword (TF-IDF) and semantic (Vector DB) search for optimal relevance.

MongoDB

Persistent storage for metadata, user profiles, and document indexes.

LLM Engine

Utilizes large language models for natural language understanding and generative Q&A.

FastAPI Server

Provides a high-performance API for all backend functionalities and user interactions.

Technology Stack: Powering the System



Our selection of technologies emphasizes open-source solutions, performance, and scalability, providing a robust foundation for the project.

USER INTERACTION

Seamless Workflow: From Upload to Answer

Upload

Users securely upload PDF documents to the system.

Process

Documents are extracted, chunked, and prepared for indexing.

Index

Text and embeddings are stored in MongoDB and the Vector Store.

Search

Users query the knowledge base using natural language.

AI Answer

LLM provides precise and contextual answers based on retrieved information.

Impact & Versatility: Advantages & Applications

Key Advantages

- **Accurate Retrieval**

  Semantic search ensures highly relevant results.

- **Rapid Response**

  Efficient processing delivers answers quickly.

- **Enhanced Privacy**

  Local processing options protect sensitive data.

Diverse Applications

- **Academic Institutions**

  Streamlining research and study for students and faculty.

- **Corporate Environments**

  Quick access to internal documentation and reports.

- **Legal Documentation**

  Efficiently navigate complex legal texts and precedents.

- **Scientific Research**

  Accelerating literature reviews and knowledge discovery.

Conclusion & Future Directions

Project Summary

We have successfully developed an AI-powered PDF knowledge base, integrating semantic search and LLM Q&A for efficient and intelligent document interaction.

Future Enhancements

- *Multi-language support for broader applicability.*
- *Development of a mobile application for on-the-go access.*
- *Advanced cloud scaling for enterprise-level deployments.*
- *Integration with external APIs and data sources.*