RV College of Engineering®
Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

RV College of Engineering®

(An autonomous institution affiliated to VTU, Belagavi)

Department of Master of Computer Applications

III SEMESTER – MCA462N Internship Project

Internship Project Title

AI Knowledge Assistant: A Cloud-Based System for Intelligent Document Query and Retrieval

Synopsis

With the rapid increase in digitized content such as research papers, manuals, and reports, users face challenges in efficiently retrieving relevant information from large collections of unstructured documents. The domains of **Cloud Computing**, **Artificial Intelligence (AI)**, and **Natural Language Processing (NLP)** offer powerful tools to address this challenge. Traditional keyword-based search systems often fail to interpret context or user intent, resulting in incomplete or irrelevant results. Recent advancements such as **Large Language Models (LLMs)** and techniques like **Retrieval-Augmented Generation (RAG)** have shown promise in providing more accurate, context-aware information retrieval. This project leverages these emerging technologies to create an AI-based system that enables users to ask natural-language questions and retrieve document-grounded answers in real time. The core objective is to develop a scalable, cloud-hosted solution that simplifies semantic querying over PDF documents.

The project methodology involves designing a modular, full-stack application using **React** for the frontend, **Flask** for the backend, and **Azure Cloud Services** for infrastructure. Key technologies include **Azure Blob Storage** (for secure PDF storage), **Azure Cognitive Search** (for semantic indexing and retrieval), and **OpenAI GPT models** (for generating human-like responses). Users authenticate via GitHub OAuth, upload their documents, and interact with a chatbot interface to pose natural language queries. The backend retrieves semantically relevant passages using Azure Search and passes them to the OpenAI model, which generates concise answers. This Retrieval-Augmented Generation workflow ensures responses are grounded in the user's actual data. The system consists of modules for file upload, document indexing, semantic search, and response generation—all integrated through secure RESTful APIs.

The expected outcome is a fully functional, multi-user platform capable of answering questions accurately based on the user's uploaded PDFs. Experimental testing indicates a high success rate in retrieving contextually relevant answers, with qualitative accuracy exceeding 90% across various datasets. The project improves accessibility to knowledge, reduces manual search effort, and enhances productivity for students and researchers. It also demonstrates the practical benefits of integrating **AI with cloud-native infrastructure**. Future extensions may include multi-format file support, vector search with Azure AI Search, and institution-level authentication via Azure Entra ID. Overall, this project showcases the effective application of modern AI and cloud technologies for solving real-world knowledge retrieval problems.

If you'd like this formatted as a printable PDF or Word document with signature lines included, just let me know!

RV College of Engineering®

(An autonomous institution affiliated to VTU, Belagavi)

Department of Master of Computer Applications

Student Signature **Guide Signature** **Director**

Note: Students already working in any CoE can continue the same work as a part of minor project under the same guide