NITTE INSTITUTE OF PROFESSIONAL EDUCATION

NITTE (Deemed to be University)

A Project Synopsis Report On

KnowCamp: An Intelligent Document-Based Information Retrieval System

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Computer Applications (BCA Honours)

Nitte (Deemed to be University)

By

Yadhu Krishnan R

NU22UCA062

Under the Guidance of

Ms. Shabari Shetty

Assistant Professor

Department of Computer Applications

Nitte Institute of Professional Education

Nitte (Deemed to be University)

Kodakal, Kannur Post

Mangalore – 575007

2025–2026

Abstract

KnowCamp is a web-based intelligent information retrieval system designed to help educational institutions manage and provide access to their official information in a simple, reliable, and transparent manner. The system allows institutions to upload official documents such as rules and regulations, academic schedules, examination timetables, circulars, and event notifications into a centralized platform. Users can then interact with the system by asking questions in natural language and receiving accurate answers generated strictly from the uploaded documents.

The core functionality of KnowCamp lies in its ability to analyze institutional documents and retrieve relevant information in response to user queries. Instead of relying on keyword-based searches, the system understands the intent of the user's question and extracts relevant content from the stored documents. The generated response is presented in a clear and concise format, along with the name of the source document used, ensuring authenticity and trustworthiness of the information provided.

The system incorporates secure user authentication and role-based access control to ensure that sensitive operations such as document uploads and administrative actions are restricted to authorized users only. This enables institutions to maintain control over their information while allowing students and staff to access verified content efficiently. By reducing the need for manual document searching and administrative intervention, the system improves information accessibility, minimizes confusion, and enhances overall communication within the institution.

KnowCamp is designed to be scalable and adaptable, allowing future extensions to support additional institutions or domain-specific deployments. The project demonstrates the effective use of modern web technologies and intelligent document processing techniques to solve real-world information access challenges in an academic environment.

Problem Statement

In the present digital era, organizations such as educational institutions and enterprises rely heavily on documents to communicate rules, policies, schedules, and important notices. These documents are commonly stored in formats such as PDFs, Word files, or text documents and are often distributed across multiple platforms. As the volume of documents grows, users face increasing difficulty in locating specific and accurate information quickly.

Most existing systems depend on manual reading or basic keyword-based search methods, which are inefficient and time-consuming. Users are often required to search through lengthy documents to find relevant details, leading to confusion and misinterpretation. In many cases, users rely on unofficial sources or repeated clarification from administrators, increasing workload and the risk of misinformation.

Another significant challenge is the lack of transparency and verification in information retrieval. Many systems provide answers without clearly indicating the source document, making it difficult for users to trust the correctness and authenticity of the information. This issue is critical when dealing with official rules, policies, or agreements.

Additionally, current document management solutions often lack proper security mechanisms such as role-based access control. Sensitive documents may be accessible to unauthorized users, while authorized users may not have appropriate access. This creates concerns related to data privacy and information security.

Therefore, there is a need for a secure and intelligent document-based information retrieval system that can provide accurate answers strictly based on uploaded documents, display the source of the information, enforce role-based access control, and reduce manual effort. The absence of such a system results in inefficiency, lack of trust, and increased administrative burden.

Objectives

- To develop an intelligent document-based information retrieval system that provides accurate answers strictly based on uploaded documents.
- To allow organizations to centrally upload and manage official documents in a secure manner.
- To enable users to ask questions in natural language and receive relevant responses efficiently.
- To implement role-based access control to ensure authorized access to documents and system features.
- To improve transparency by displaying the source document used for generating each answer.
- To reduce manual effort and time spent searching through lengthy documents.
- To design the system in a scalable manner so that it can be extended to other organizations or domains in the future.

Methodology

- The system begins with user authentication, where users log in securely and are assigned roles such as administrator, staff, or user.
- Role-based access control is applied to ensure that only authorized users can upload, manage, or access specific documents.
- Authorized users upload documents such as PDFs or text files, which are stored securely along with relevant metadata.
- Uploaded documents are processed and converted into searchable content to enable efficient information retrieval.
- When a user submits a question, the system analyzes the query and retrieves relevant sections from the uploaded documents.
- An intelligent response is generated strictly based on the retrieved document content, without adding external information.
- The system displays the generated answer along with the source document name to ensure transparency and trust.
- All interactions are handled through a web-based interface for easy access and usability.

Expected Outcomes

- Development of a reliable web-based application that allows users to retrieve information directly from uploaded documents.
- Improved efficiency in accessing institutional information by enabling users to ask questions in natural language.
- Accurate and consistent answers generated strictly from official documents, reducing ambiguity and misinformation.
- Increased transparency by clearly displaying the source document used for each response.
- Secure handling of data through proper authentication and role-based access control mechanisms.
- Reduced dependency on manual communication with administrative staff for routine information queries.
- Better document management and organization within institutions through centralized storage.
- A flexible and scalable system architecture that can be enhanced to support additional features and broader use cases in the future.

Future Scopes

- Extension of the system to support multiple institutions within a single platform.
- Integration with additional domains such as legal, healthcare, and corporate policy management.
- Support for more document formats and multilingual content.
- Automatic updates and version control for newly uploaded or revised documents.
- Enhanced analytics to track frequently asked questions and user interactions.
- Mobile application support for improved accessibility.
- Integration with external systems such as institutional portals and notification services.
- Advanced security features to further strengthen data protection and access control.

References

- FastAPI Documentation, FastAPI: Modern Web Framework for Python, Available at: https://fastapi.tiangolo.com
- React Documentation, React – A JavaScript Library for Building User Interfaces,Available at: https://react.dev
- PostgreSQL Documentation, PostgreSQL Official Documentation, Available at: https://www.postgresql.org/docs
- Lewis, P., et al., Retrieval-Augmented Generation for Knowledge-Intensive NLPTasks, arXiv preprint, 2020.
- National Institute of Standards and Technology (NIST), Role-Based Access Control(RBAC) Model.
- Jurafsky, D. and Martin, J.H., Speech and Language Processing, Pearson Education.
- Python Software Foundation, Python Programming Language Documentation, https://docs.python.org