Synopsis

Title : DocSeek

Aim : The goal is to develop a web application that can analyse different kinds of content and give users a useful outcome.

Description: Our Lang Chain software DocSeek is made to evaluate various kinds of documents and reply to user queries. Through the application of strong natural language processing algorithms, the app responds correctly to inquiries from users. The platform improves user interaction and information usage by constantly processing a range of input forms.

Modules:

Login Module: This model provides safe access to the web application and manages identities of users. It limits access to private features and data, maintains user sessions, and verifies login details.

Translation Module: Text conversion through languages is made simple by the translation model. By translating user input and responses automatically, it improves multiple language communication.

Doc Insight : It allows users to interact with document files that they have submitted. In order to answer user queries, it processes and analyses document content by extracting useful information. The model allows for document extraction and processing, which improves the web application's functionality and effectiveness.

CSV Explorer: The Interactions with uploaded CSV (comma-separated values) files are managed via the CSV interaction model. On the basis of the dataset, it evaluates CSV data, manipulates data, and answers questions from users. The concept enables users to derive conclusions from CSV files within the chat interface application by handling tabular data with ease.

Synopsis

Hardware Specifications:

Operating System: Windows 7 or 10

RAM : Minimum 4gb up to 8gb is better.

Web Browser : Google Chrome, Mozilla Fire-Fox or any compatible browser.

Hard Disk : Minimum 20gb.

Software specifications:-

Python: which includes pip (Pip Installs Packages), to manage dependencies and run Python based build tools and libraries.

Development Tools: API development tool (Open-AI API , Huggingfacehub API), Code-Editor (VS Code).

Front-end (Client-side):

Stream-lit : A python library used to create custom web app.

Back-end (Server-side):

Python: Python has a huge and active library and package system that makes handling HTTP requests and backend development tasks easier.

Python Libraries :

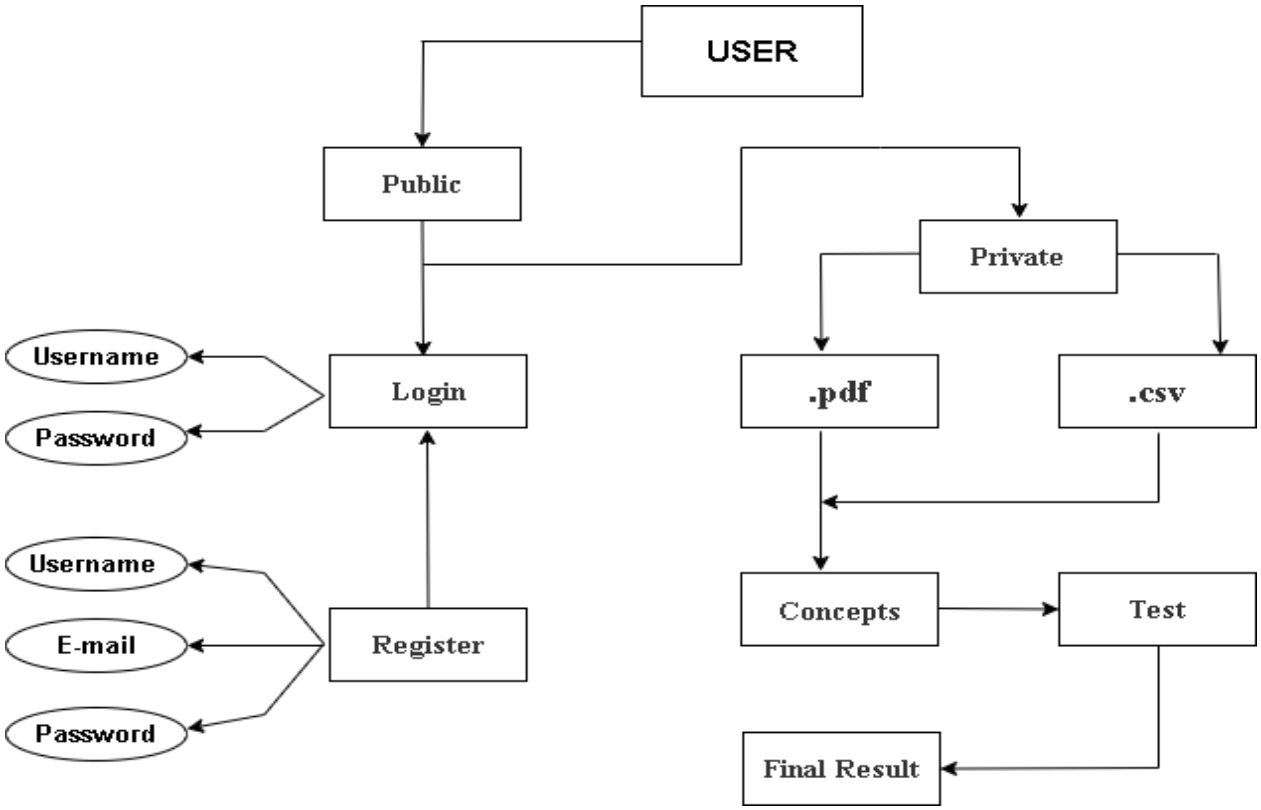
NumPy : It is a basic computing toolkit that supports matrices, arrays, and mathematical functions.

Pandas : A powerful library for data analysis and manipulation that offers data structures like Data Frame.

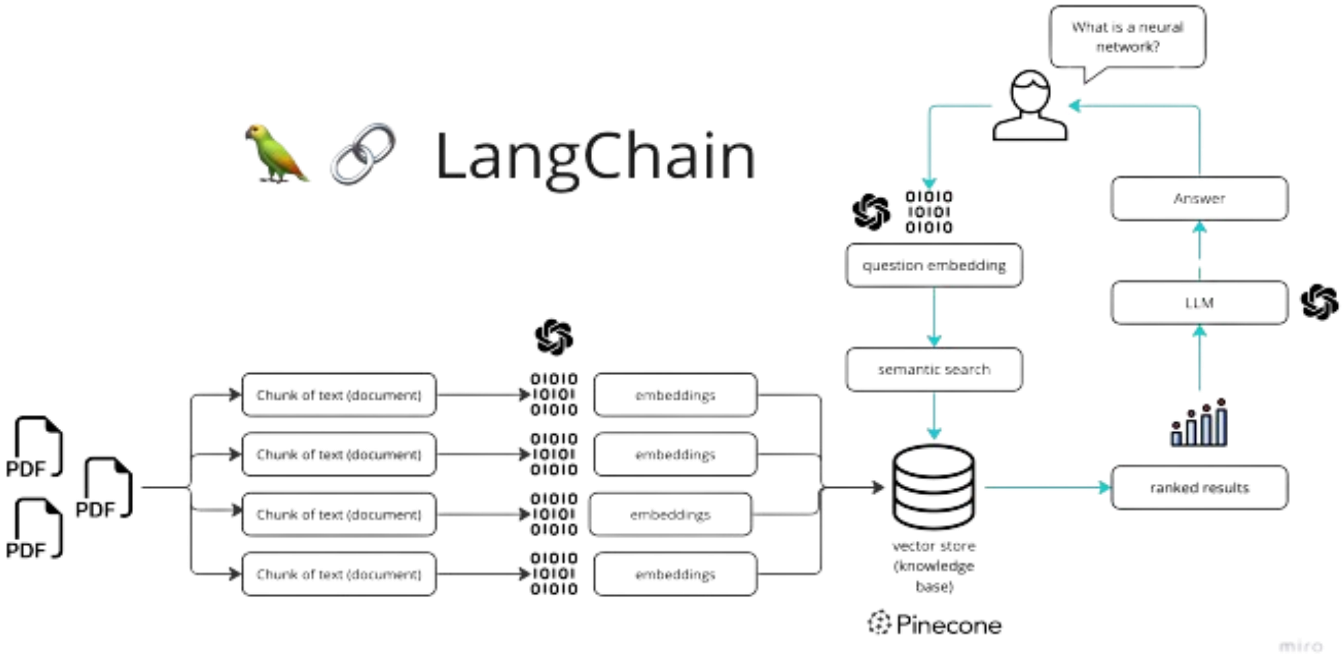
Lang Chain: A framework called Lang Chain is used to create applications that are based on language models. It allows applications that rely on a language model to reason (about how to answer based on provided information, what actions to take, etc.) and connect a language model to sources of context (prompt instructions, quick brief samples, content to locate the answer in, etc.).

Synopsis

ER-Diagram :



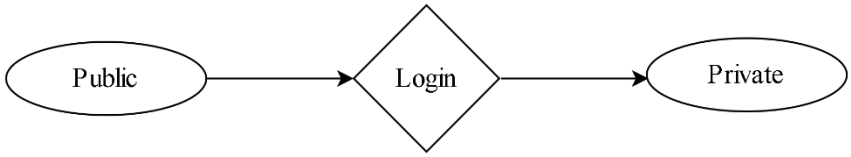
UML Diagram :



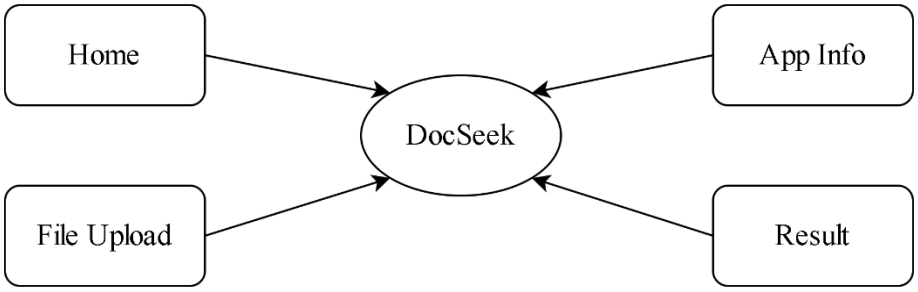
Synopsis

Data Flow Diagram :

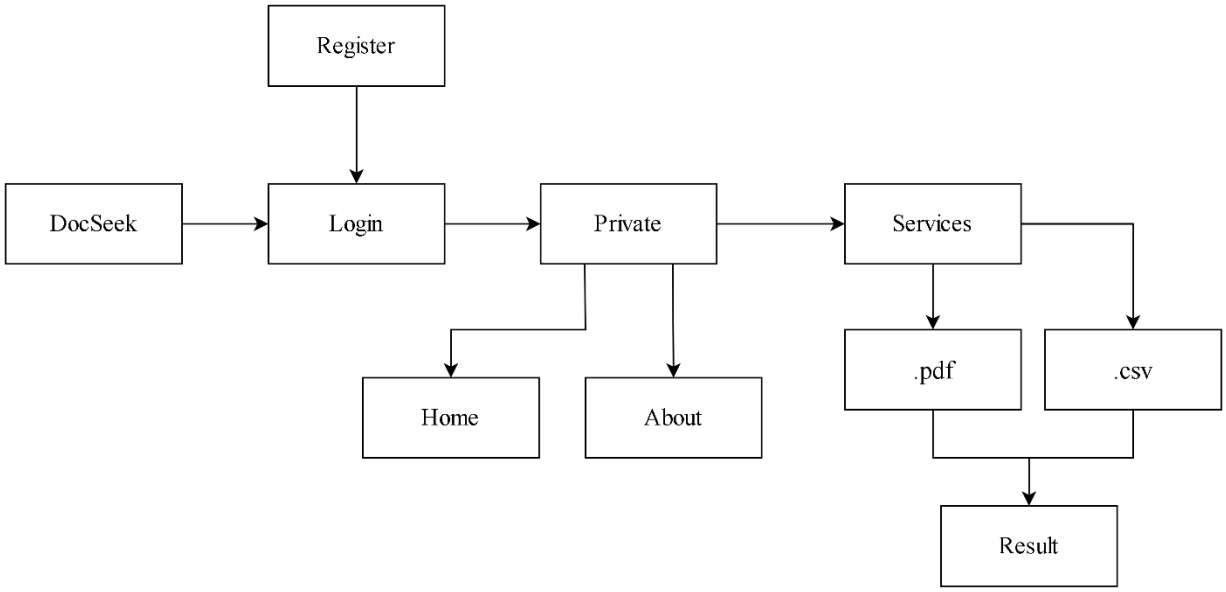
Level Zero DFD :



Level One DFD :



Level Two DFD :



Synopsis

Conclusion:

Coming back up, the web application offers consumers a complete solution for intelligent analysis of CSV and PDF data. Natural language processing, relationship for several formats, expert search options, and interaction with external services are just a few of the application's many features. It promises to deliver useful data quickly and effectively. In order to provide a user-friendly experience, the project focuses user authorization and scalability.

Future Score:

Natural Language Understanding (NLU) Improvements: Enhancements to Natural Language Understanding (NLU): Integrate current NLU models to improve the app's language understanding skills. To provide more context-aware replies, use Named Entity Recognition (NER) to extract items from PDFs and CSV files, such as dates, locations, organizations, etc.

Support for More File Formats: Include support for other popular document formats, such as Word documents (docx), PowerPoint presentations (pptx), text files (txt), etc., in addition to PDFs and CSVs.

Integration with External APIs and Services: To enable their huge NLP skills, integrate with third-party APIs or services like AWS Comprehend, Google Cloud Natural Language API, etc.