Document Retrieval System using RAG

In propose work we are utilizing cloud services and Large Language Model called RAG (Retrieval-Augmented Generation) for efficient data retrieval. RAG model can be used to search document with high accuracy and can be utilized to generate text. Existing techniques will not utilize entire NLP vocabulary and other processing techniques like stemming, stop words removal, document weightage, lemmatization which will affect accurate document retrieval and may reduce accuracy.

Propose RAG model has inbuilt support for all NLP processing techniques and can generate accurate tokenization for input text and for searching documents which will help in accurate and efficient document retrieval.

In propose application utilizing cloud services to manage and store all official or users oceans of documents and then employing RAG model for document retrieval and for text generation.

Note: for text generation RAG required heavy models and those models required huge GB of RAM and hard disk for storage and to avoid this we have used simple model for text generation.

To implement this project we have designed following modules

1. New User Signup: user can sign up with the application
2. User Login: user can login to system
3. Upload Document to cloud: user can upload desired document which will saved in cloud memory space
4. RAG Document Retrieval: in this module user can enter some queries and then RAG model will search that query in all documents and then returned top matching documents with accuracy score
5. RAG Text Generation: using this module user can input some sentence and then RAG will generated text based on given sentence.

Note: RAG text generation may take 2 to 6 minutes of time to give you output.

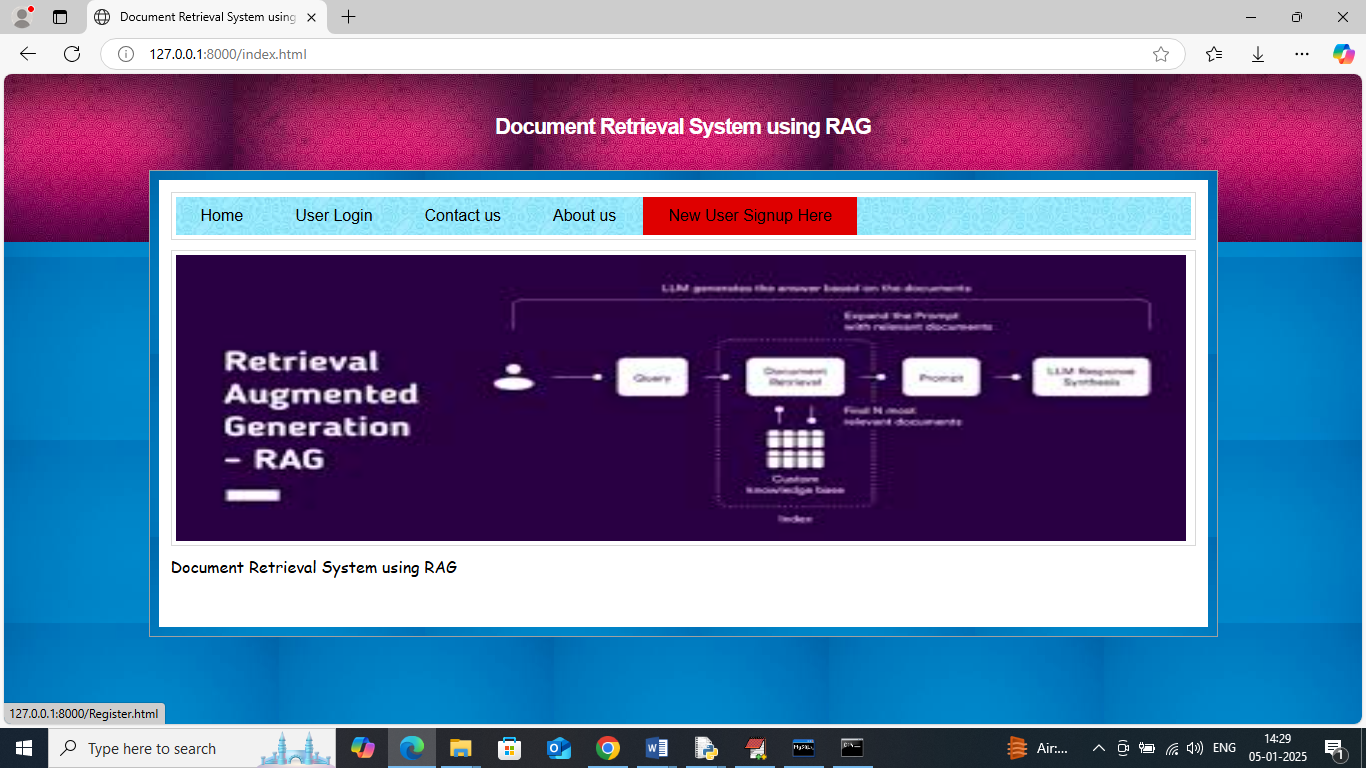
SCREEN SHOTS

To run project install python 3.7.2 and then install all packages given in requirements.txt file and then install MYSQL and then open MYSQL console and then copy content from ‘database.txt’ file and paste in MYSQL console to create database.

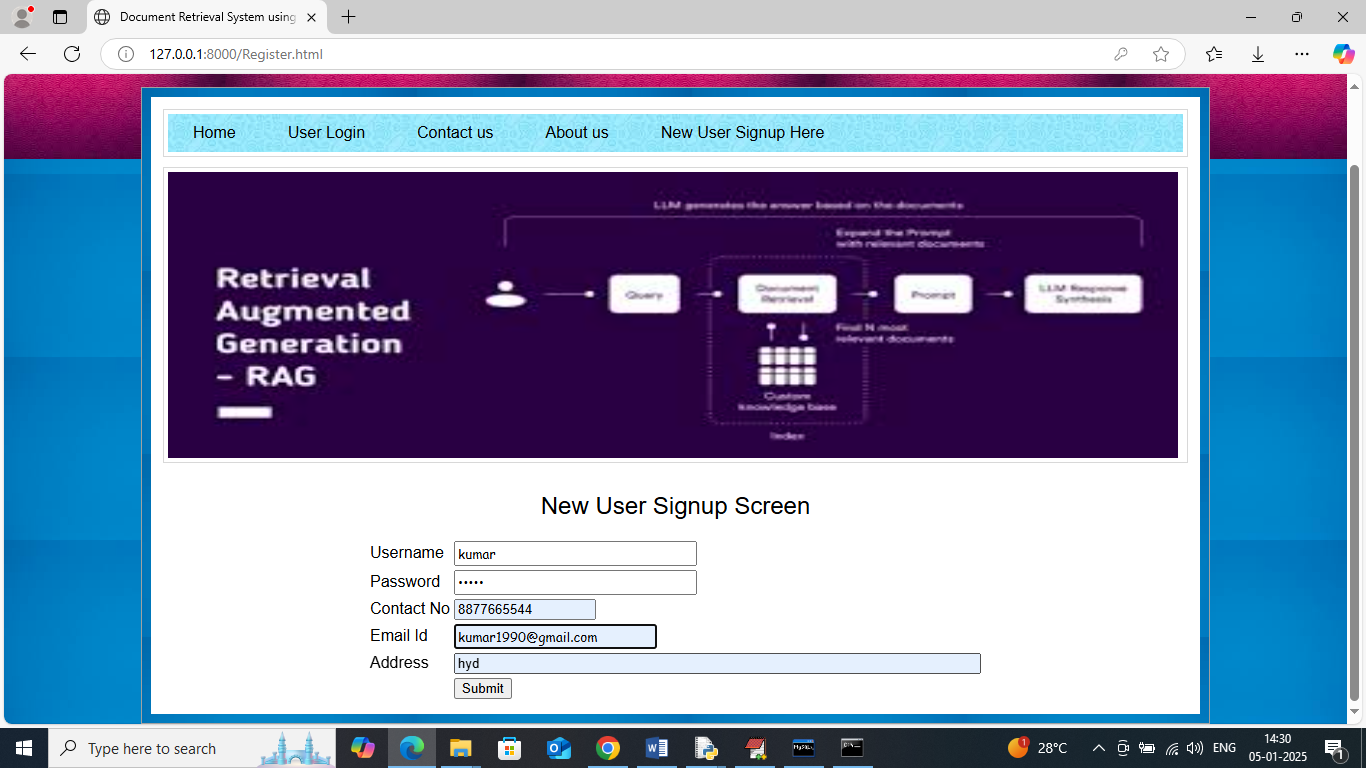
Now double click on ‘run.bat’ file to start python server and get below page



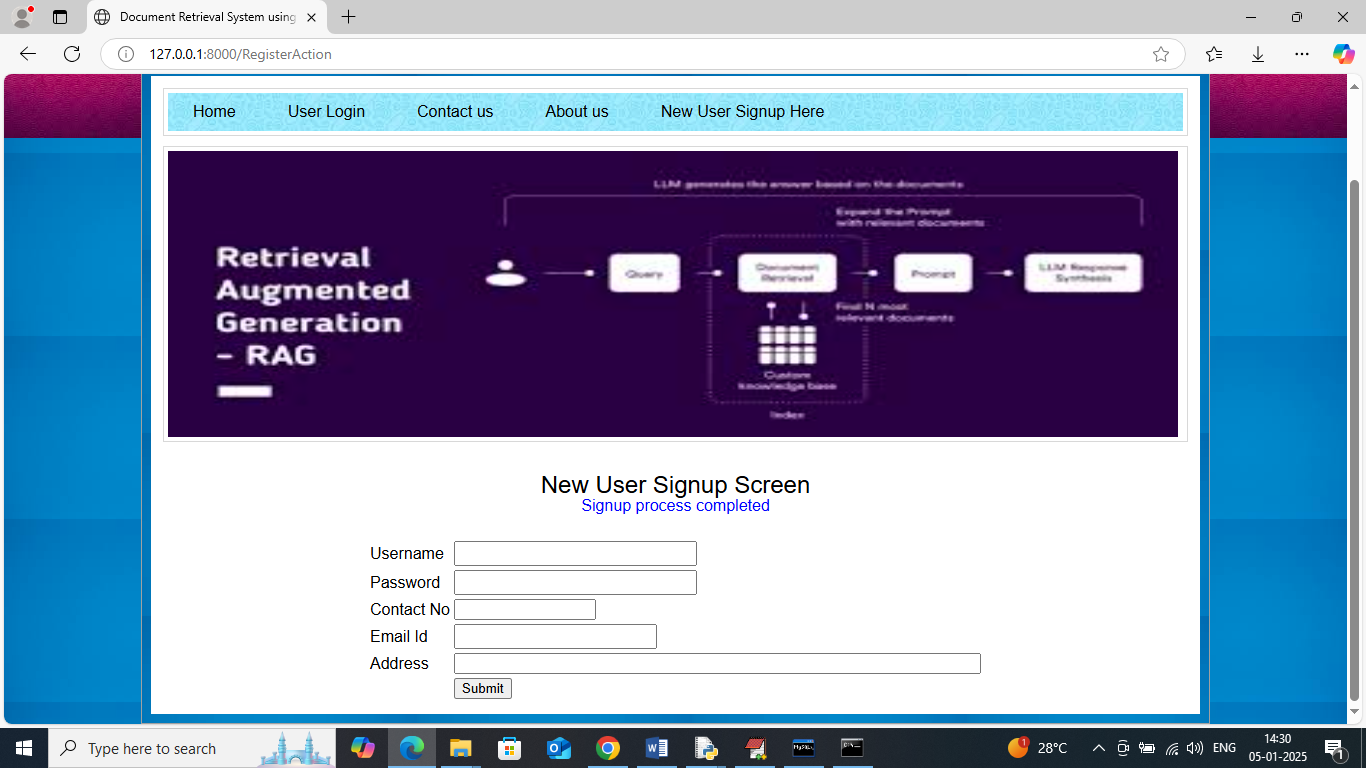
In above screen python server started and now open browser and enter URL as <http://127.0.0.1:8000/index.html> and then press enter key to get below page



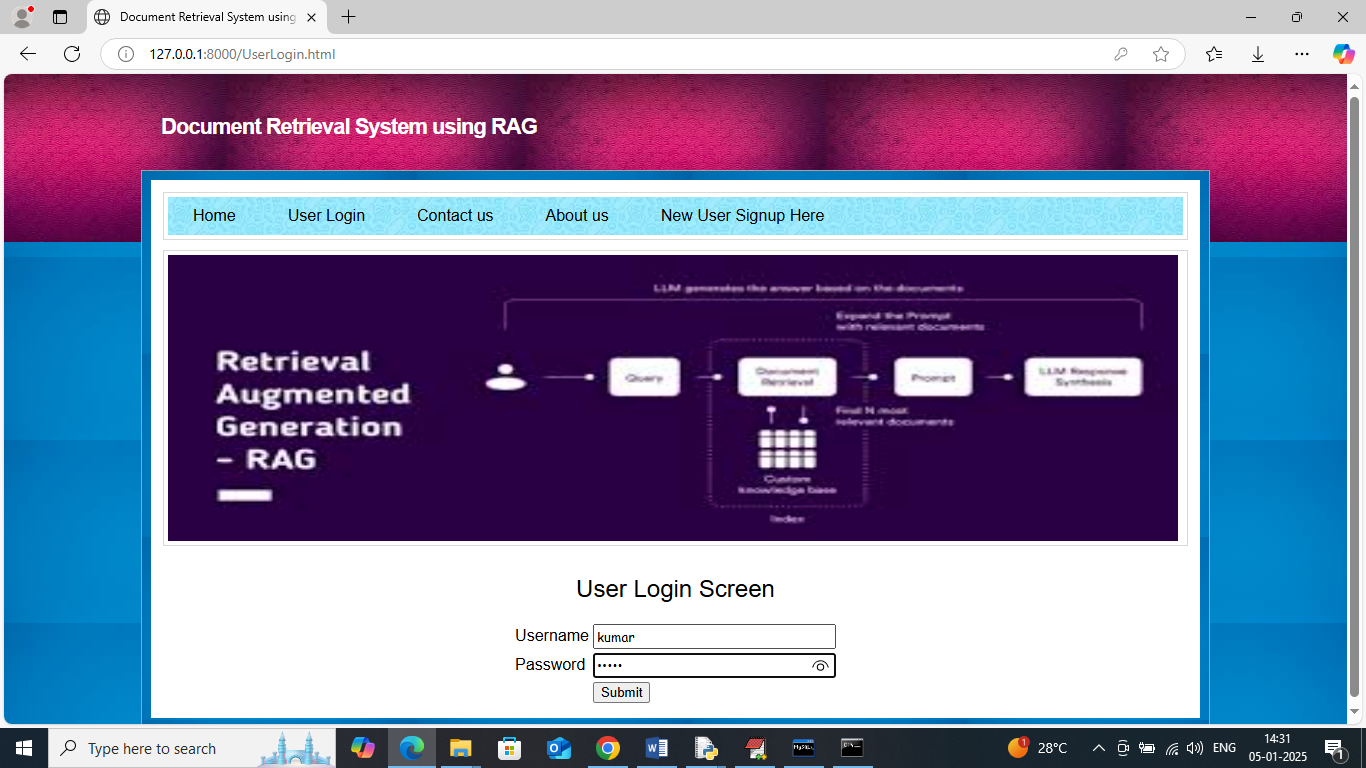
In above screen click on ‘New User Sign up Here’ link to get below page



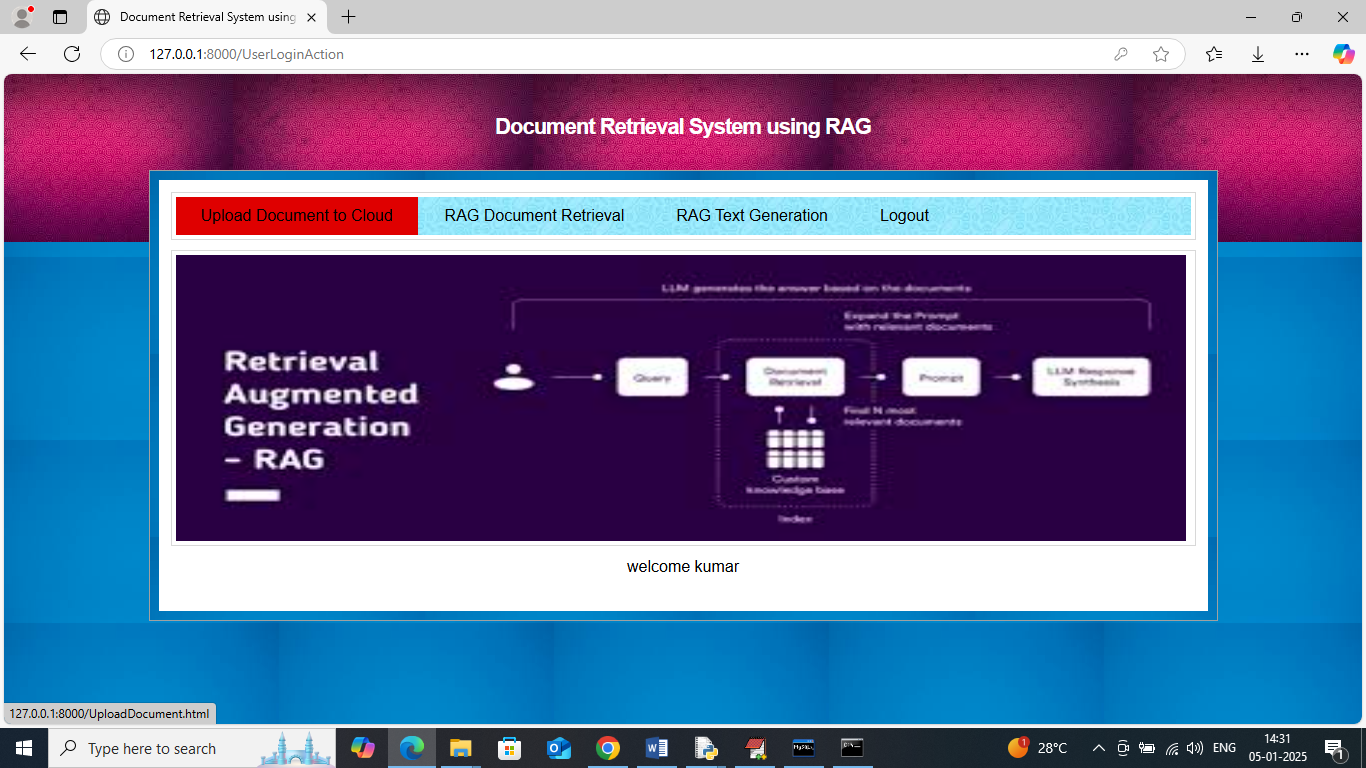
In above screen user is entering sign up details and then press button to get below page



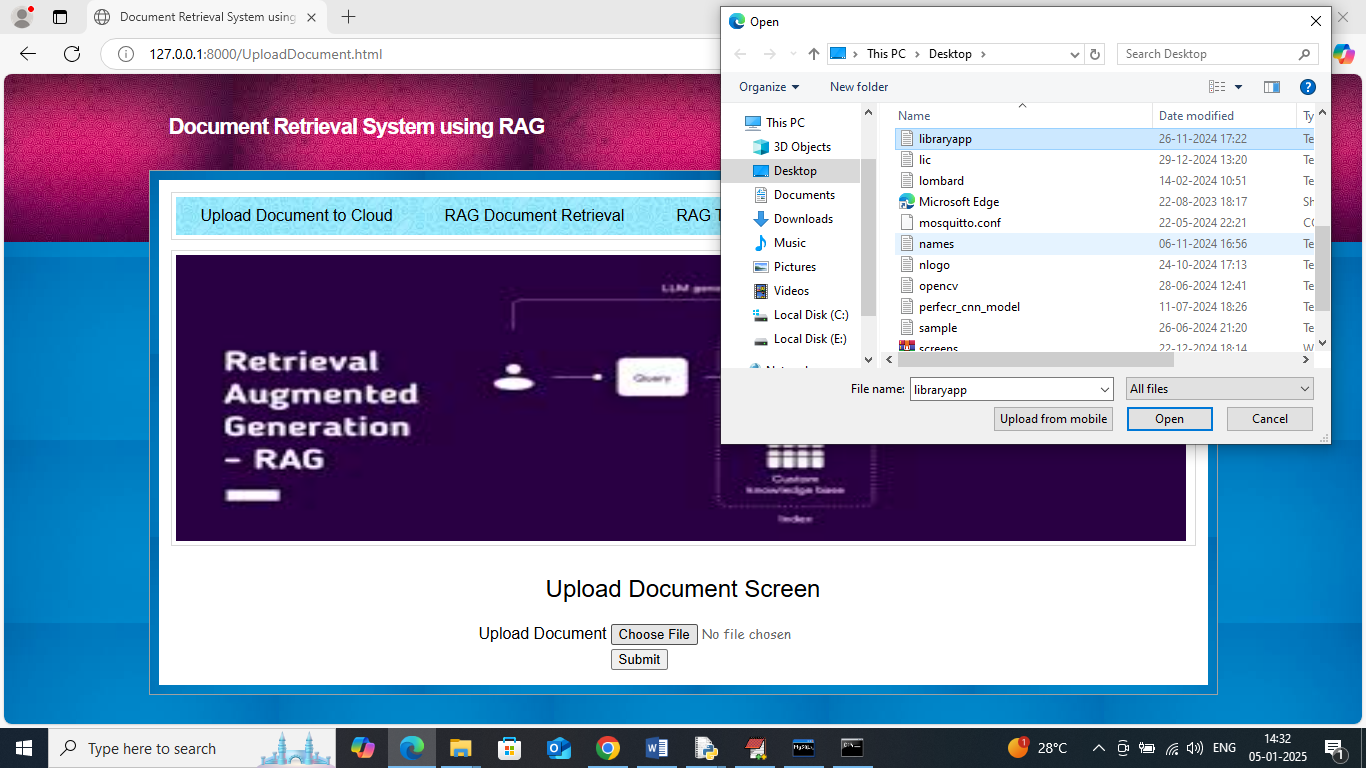
In above screen user sign up process completed and now click on ‘User Login’ link to get below page



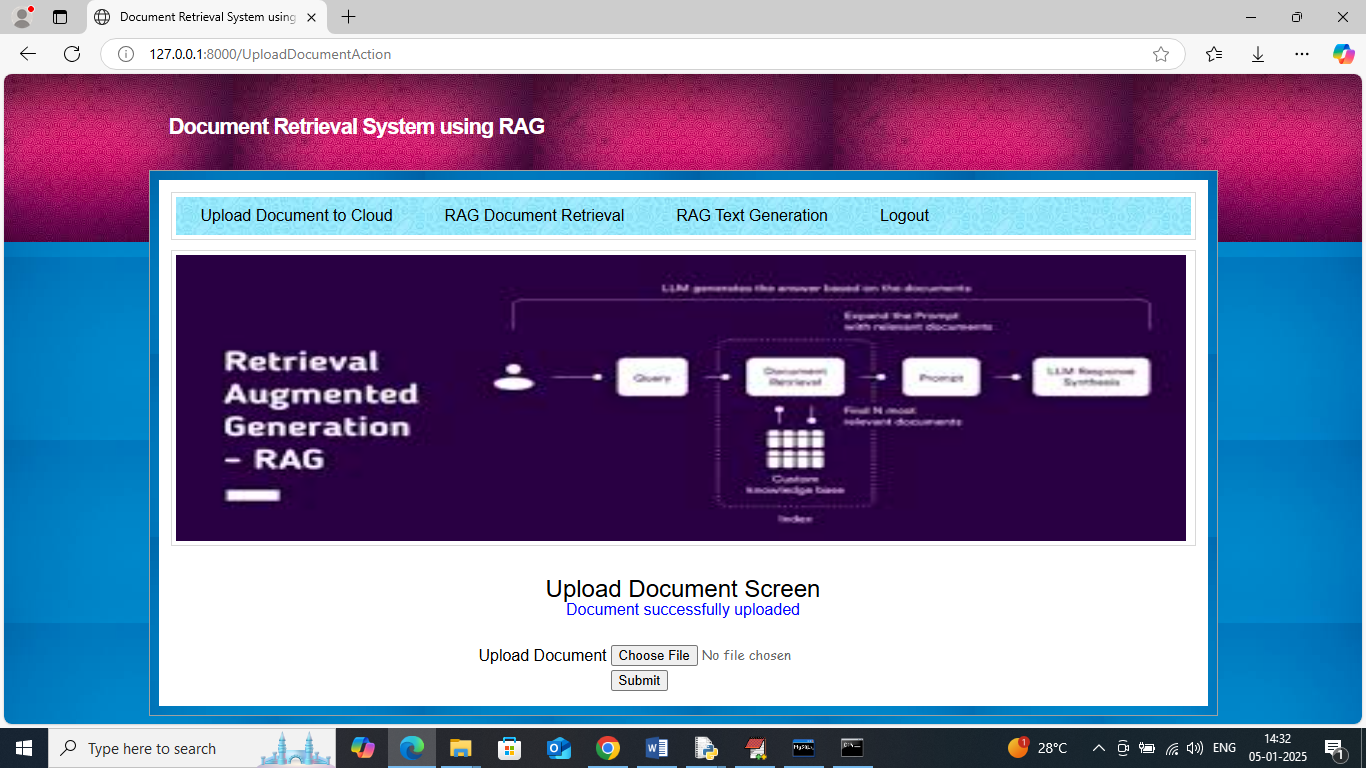
In above screen user is login and after login will get below page



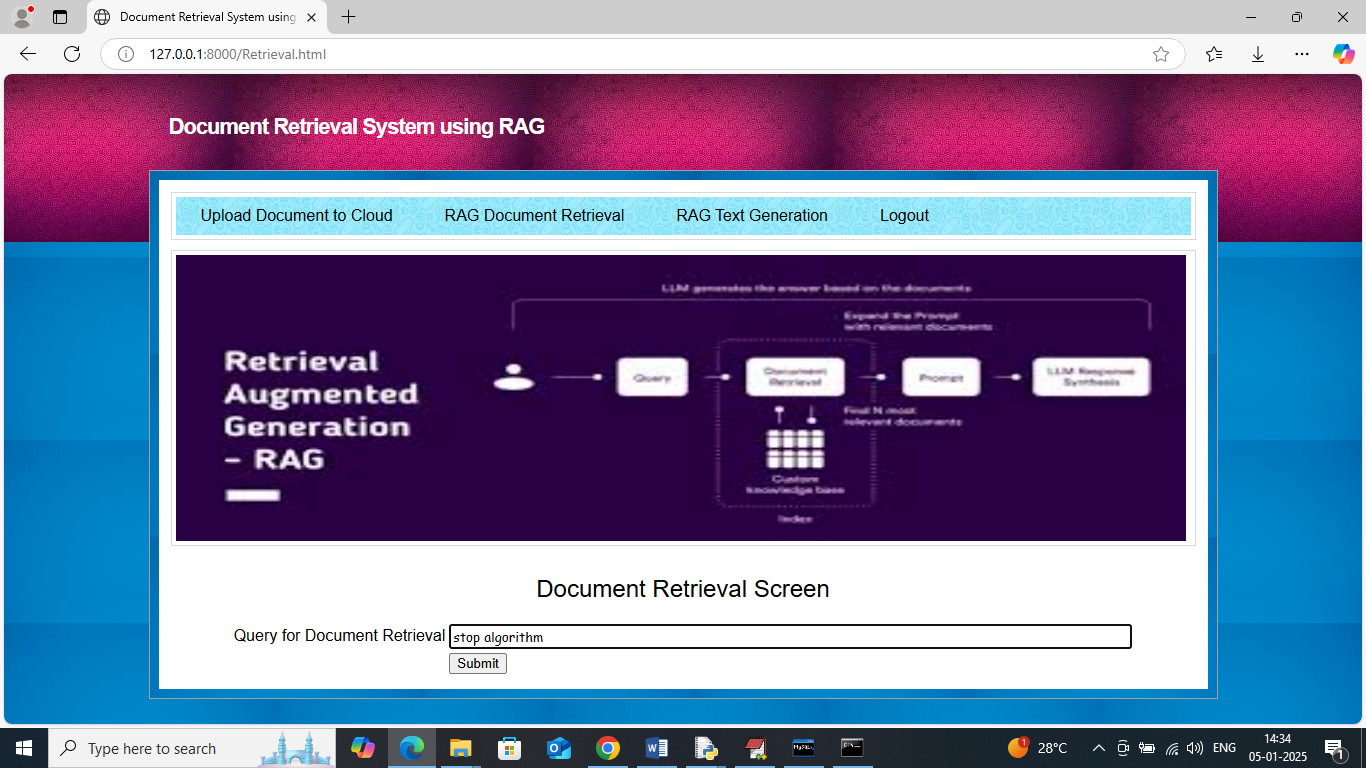
In above screen user can click on ‘Upload Document to Cloud’ link to get below page



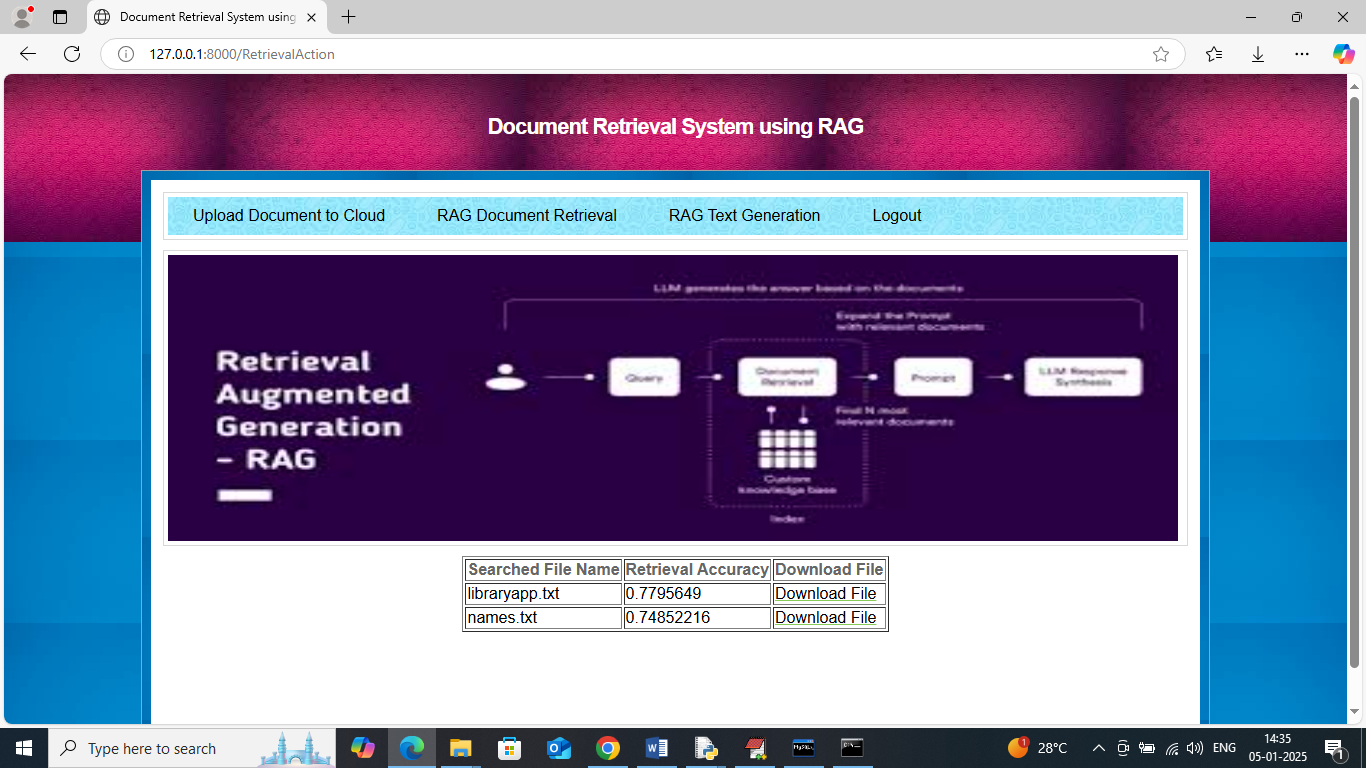
In above screen selecting and uploading text document and then click on ‘Open and submit’ button to upload document and get below page



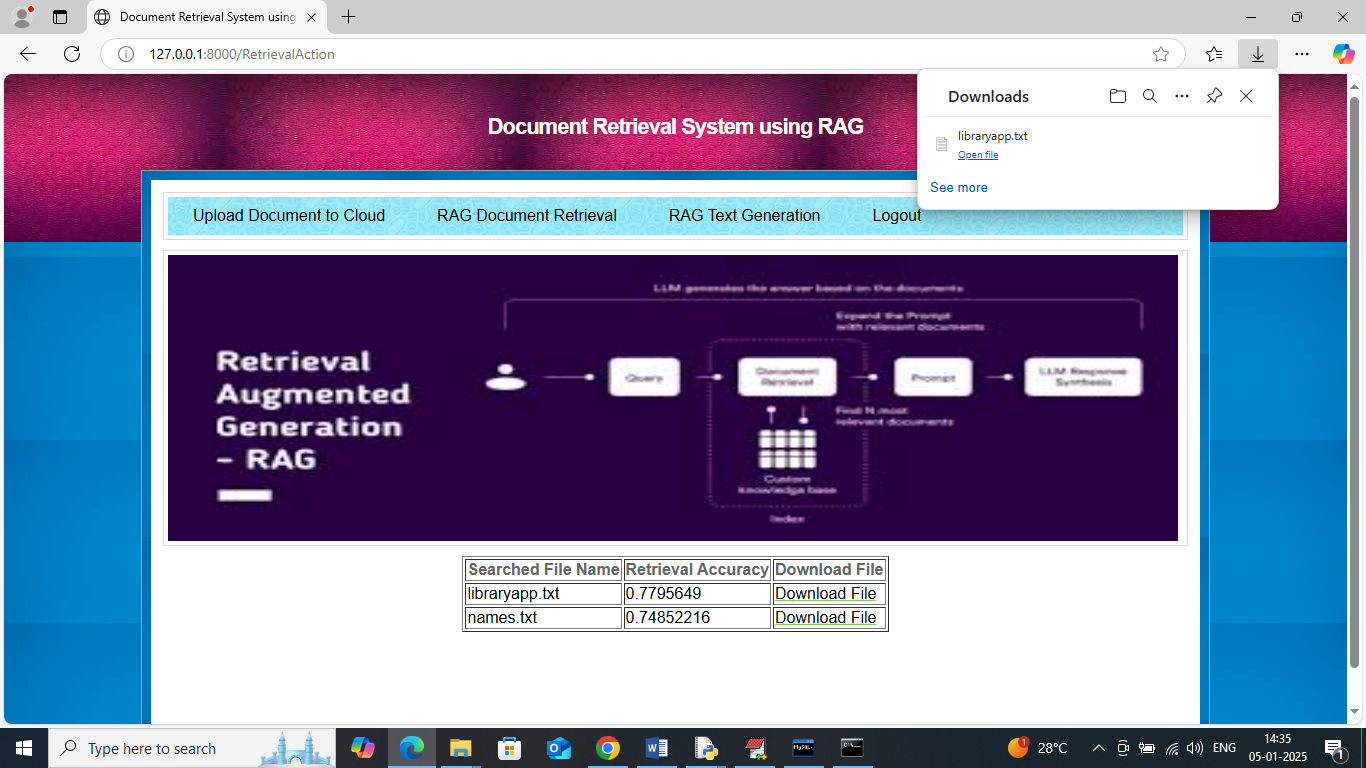
In above screen document successfully uploaded to cloud and similarly you can upload as many documents as you want and now click on ‘RAG Document Retrieval’ link to get below page



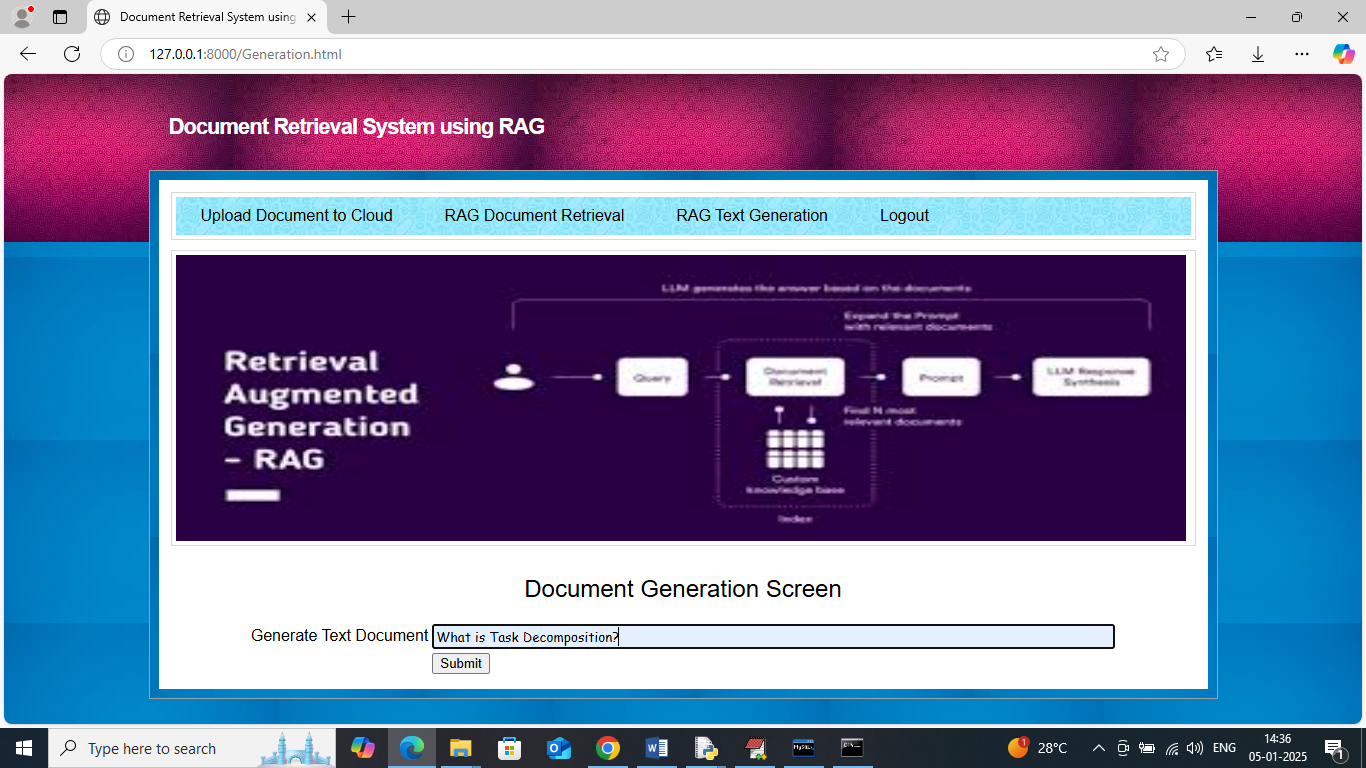
In above screen enter some query to search and retrieve documents and then will get below page



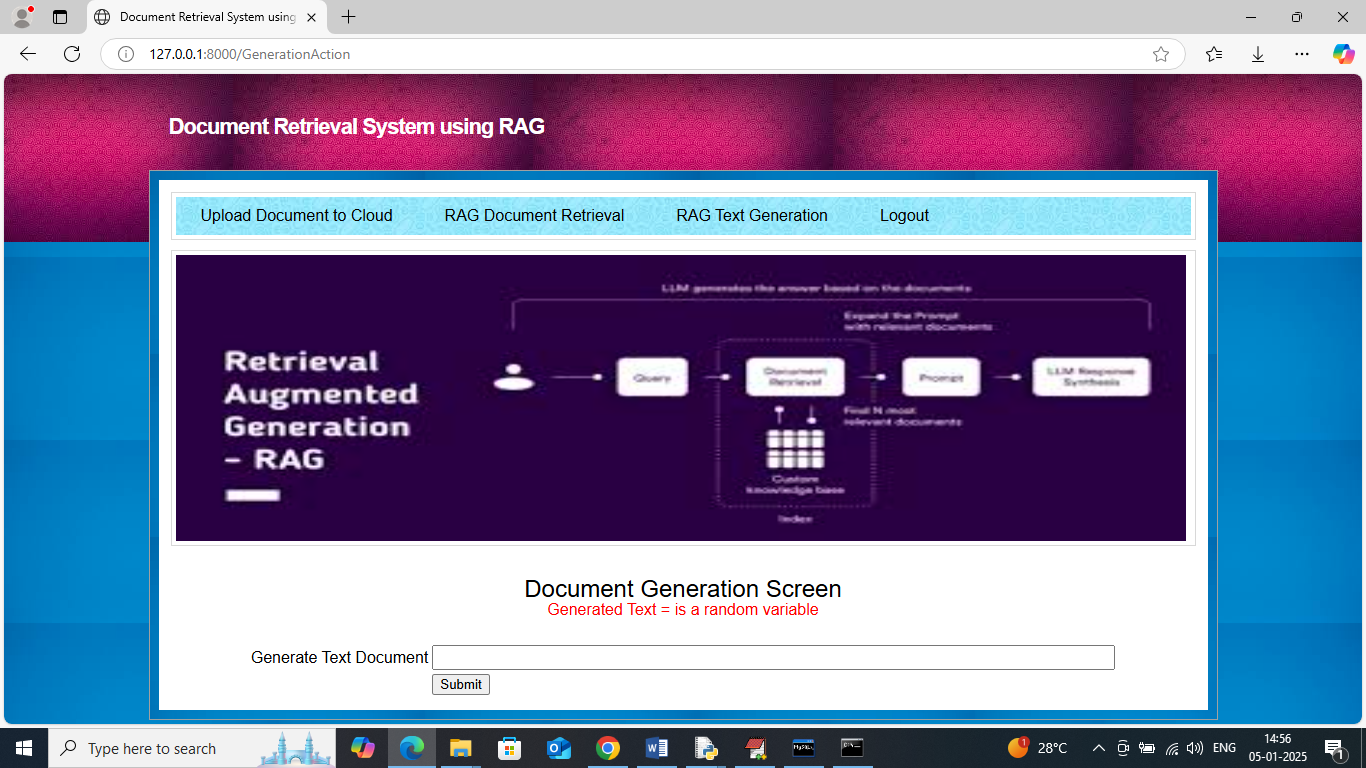
In above screen can see names of document along with retrieval accuracy and can click on ‘Download File’ link to download desired document and get below page



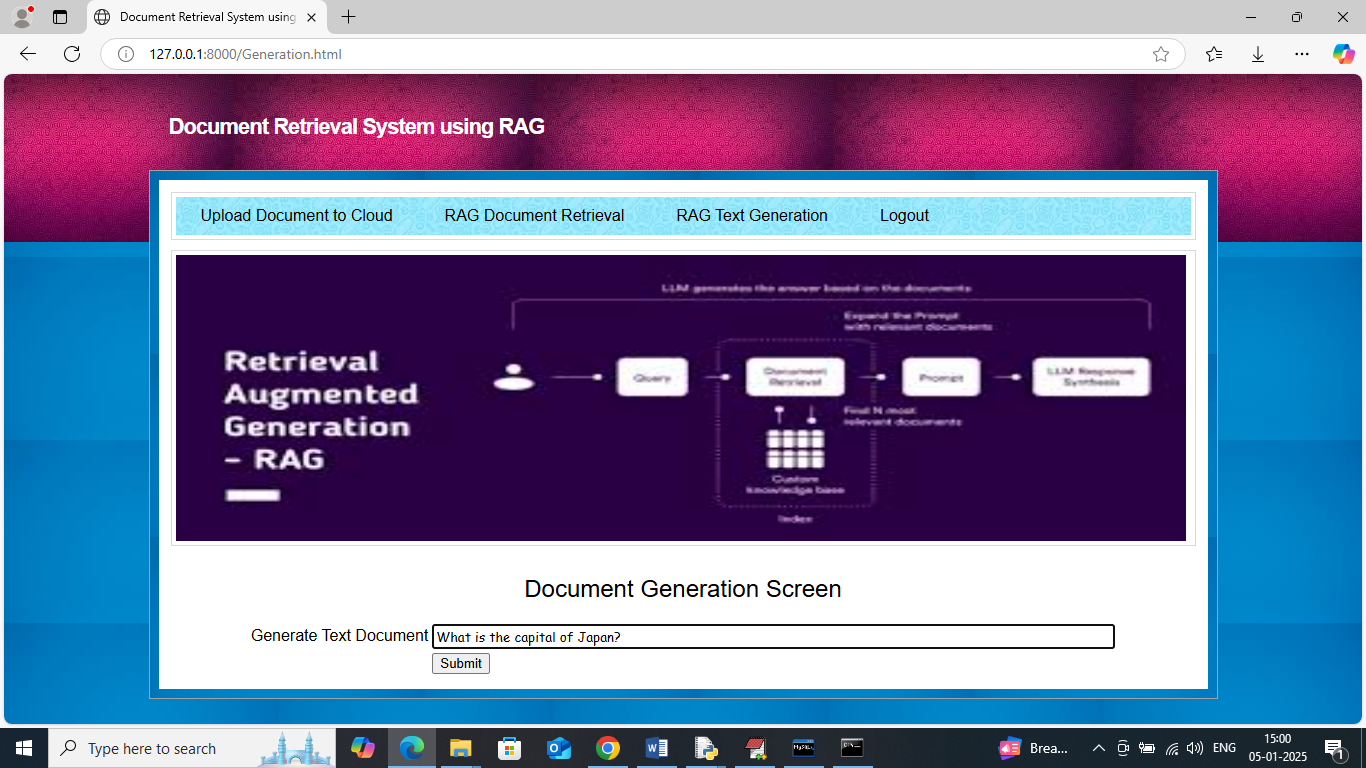
In above screen in browser status bar can see file downloaded and similarly you can search for any query and now click on ‘RAG Text Generation’ link to get below page



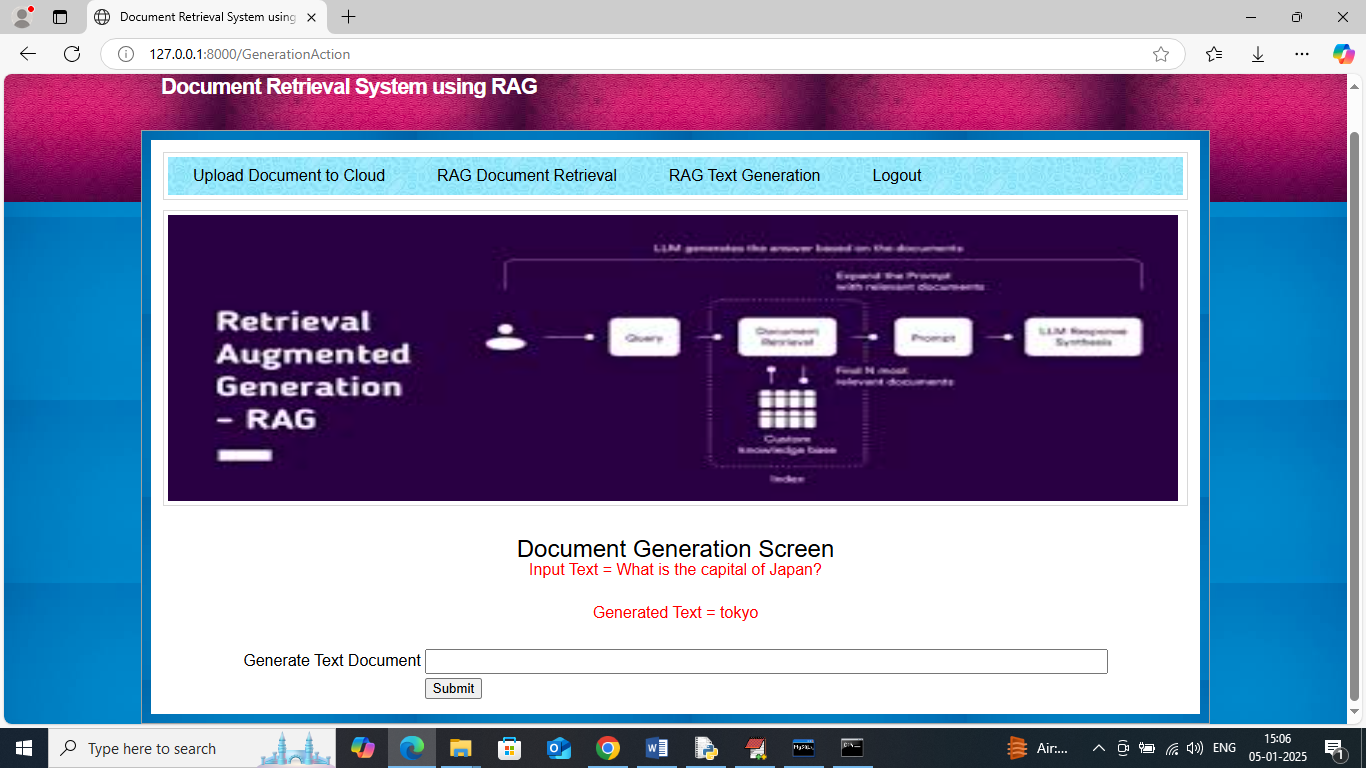
In above screen to generate text I gave some question and below is the output



In above screen for given we got generated text in red colour text and below is the another example



In above screen entered some other text and below is the output



In above screen can see output for given text question and similarly you can ask any text to generate