HEXAGON

Internship Completion Report

Name: Prayag Chawla

College: Thapar Institute of Engineering and Technology

Internship Start Date: June 3, 2024

Internship Completion Date: August 2, 2024

Course: B.E Electronics and Computer

Development of a context-aware NLP system that can understand and retrieve geospatial data based on the context provided in user queries. The system is able to infer implicit information and provide relevant geospatial data accordingly, along with a specific use case of Mars/Moon Crater Detection.

SOLUTION

1) Local Image Dataset Management:
- Enabled users to upload and process images along with associated metadata.
- Stored images and their metadata in a database for efficient querying and retrieval.

2) Article Management:
- Allowed users to add multiple articles by providing URLs.
- Fetched, parsed, and stored article content for subsequent use in answering user queries.

3) Question Answering:
- Utilized natural language processing techniques to answer user queries based on the content of stored articles.
- Provided an option to generate and download a PDF report containing the question and its answer.

4) Image Querying:
- Implemented functionality to search for images based on specific criteria such as metadata attributes.
- Displayed the relevant images and associated metadata based on the search query.

5) Feedback Collection:
- Allowed users to submit feedback through a dedicated interface.
- Generated a PDF report of the feedback and provided options to download the report.
- Integrated functionality to redirect users to their email client with the feedback pre-filled, enabling them to send the feedback directly.

TOOLS AND TECHNOLOGIES

- Python
- Streamlit: For building the interactive web interface.
- Requests: For fetching article content from URLs.
- BeautifulSoup: For parsing HTML content of articles.
- SQLite: For managing and querying the database of images and articles.
- Pandas: For handling CSV files and data processing.
- Transformers (Hugging Face): For question-answering functionality.
- FPDF: For generating PDF reports.
- PIL (Pillow): For handling and displaying images.
- NumPy: For numerical operations and similarity calculations.
- Scikit-learn: For vectorizing text and calculating cosine similarity.
- HTML, CSS

APPROACH

1) Database Setup:
- Designed and implemented a database schema to store images, metadata, and articles efficiently.
- Set up the necessary tables and relationships to support the required functionalities.

2) Image and Metadata Management:
- Developed functionality to upload and store images along with their associated metadata.
- Ensured accurate and efficient storage and retrieval of image data.

3) Article Management:
- Implemented mechanisms to fetch and store article content from provided URLs.
- Processed and prepared text data for querying and analysis.

4) Question Answering:
- Applied natural language processing techniques to answer user queries based on stored article content.
- Generated and provided options for users to download PDF reports containing query and response details.

5) Image Querying:
- Developed search functionality to retrieve images based on specific metadata criteria.
- Displayed relevant images and metadata to users based on their search queries.

6) Feedback Collection:
- Implemented a feedback collection system with options for generating and downloading feedback reports in PDF format.
- Provided a seamless redirection to email clients for sending feedback, with pre-filled content for user convenience.

IMAGES OF THE TOOL
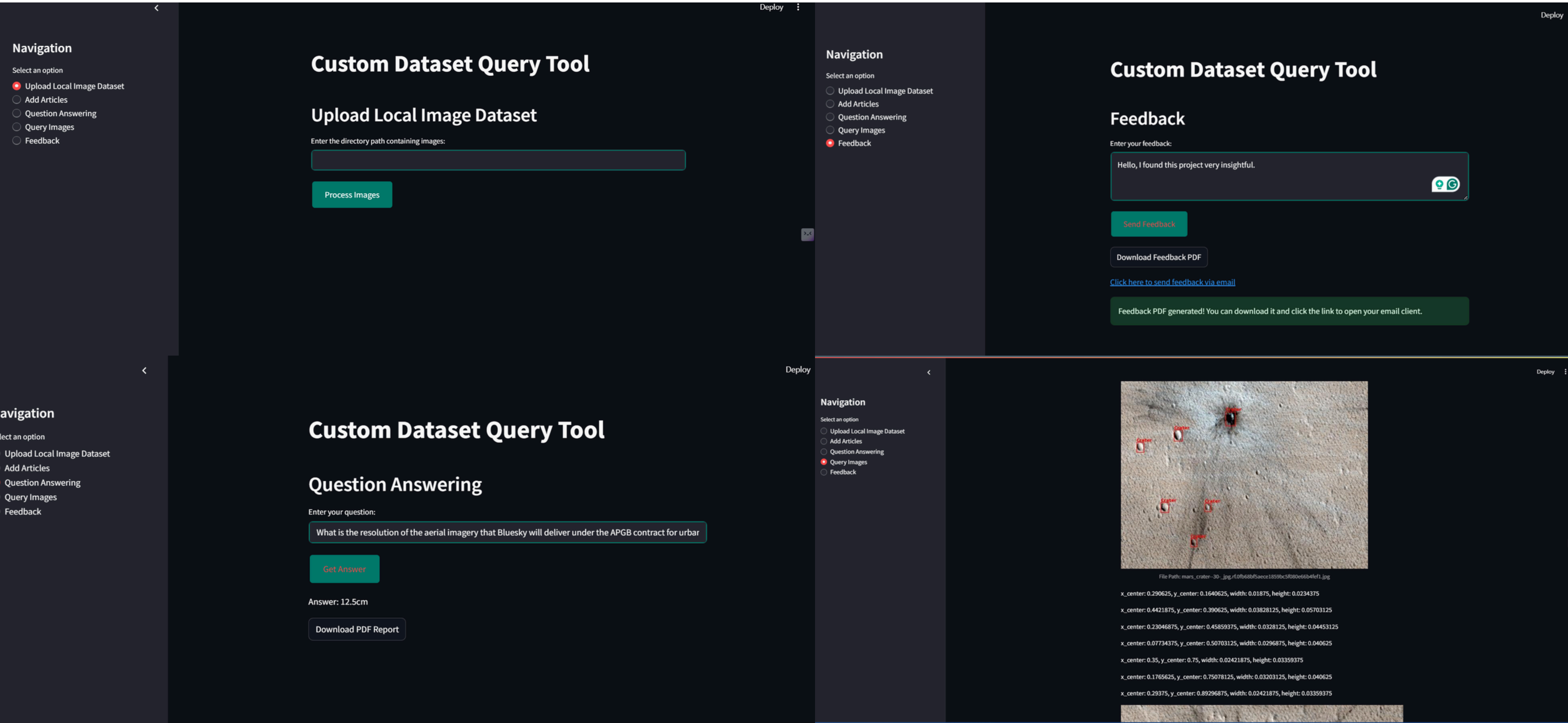
REPOSITORY LINK & SUBMISSION DETAILS

Github Link

Submitted to:

**Nitin Diwedi
Solution Architect
Hexagon India**