Metadata Project

**Project Roadmap: Metadata Search Engine (MetaHub DB)**

**Project Overview:**

Our team, consisting of three members, undertook the task of building a metadata search engine as part of the for the Metadata course. The primary objective of the project was to create a search engine capable of processing various file types (initially focusing on photos), extracting metadata, and storing it in a MySQL database. Additionally, the search engine was envisioned to handle web page URLs, extracting metadata from web pages for future scalability.   
   
**User stories**

New user stories may be added or old ones removed by the product owner, but these are the ones we currently have:

1. As a system owner, I want a MySQL database that can effectively store metadata about various types of files so that I can search for a file through a metadata search.
2. As a system owner, I want a system that can extract metadata from folders of files and save the metadata related to file names/paths in the database.
3. As a system owner, I want a MySQL database that can effectively store metadata about web pages/URLs so that I can search for a webpage through a metadata search. (Hold off on this user story until a review of scraping/harvesting data from web pages is done in week 49.)
4. As a visitor, I want a web-based interface for searching metadata so that it is easy for me to search and view results.
5. As a visitor, I want to be able to choose what I want to search (a specific file type or web pages) so that the system understands the type of data I want to search for.
6. As a visitor, I want to be able to search both by file name and metadata so that I can easily find what I'm looking for.
7. As a visitor, I want to be able to search effectively in the metadata (e.g., if something is equal to, not equal to, greater than, less than) a certain value, so that it is easy for me to refine my search.
8. As a visitor, I want to be able to search geographically when the metadata contains geographic information (latitude and longitude).
9. As a visitor and system owner, I want the system to have a set of test data (non-copyright protected) so that I can test that the system functions as intended.

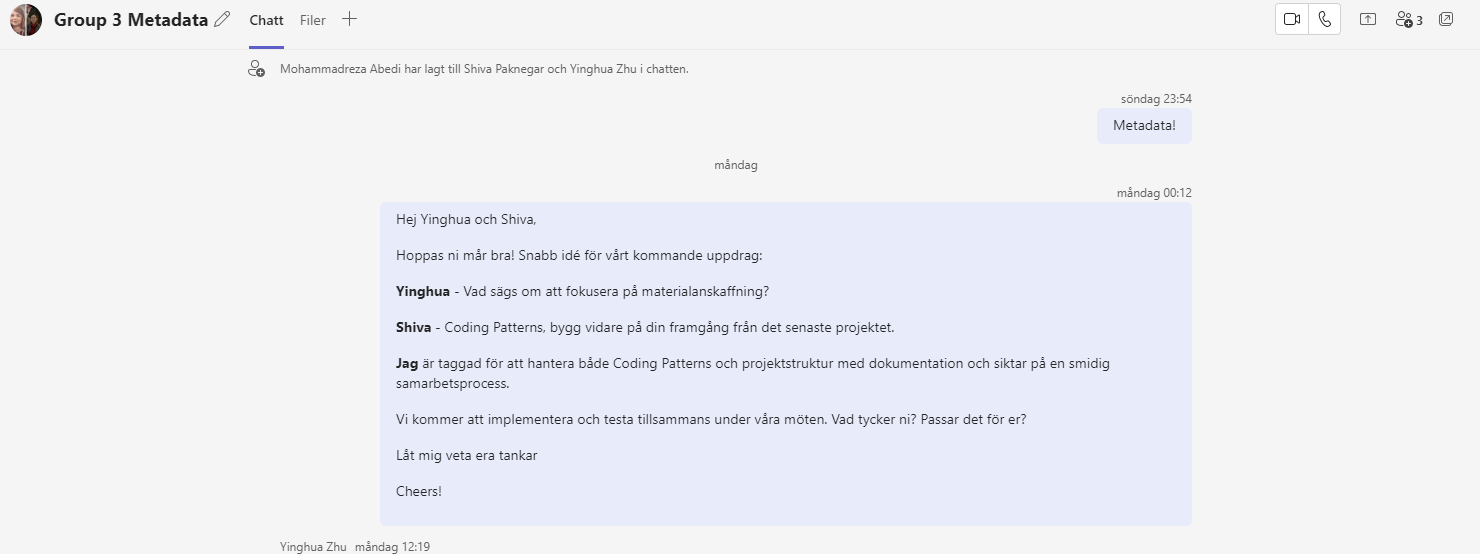
**Technologies Used:**

* **Node.js:** Employed for metadata extraction from files and web pages using relevant libraries and plugins.
* **MySQL:** Chosen as the cloud-based server for storing extracted metadata.
* **Git (GitHub):** Utilized for version control to facilitate collaborative development.

**Team Collaboration:**

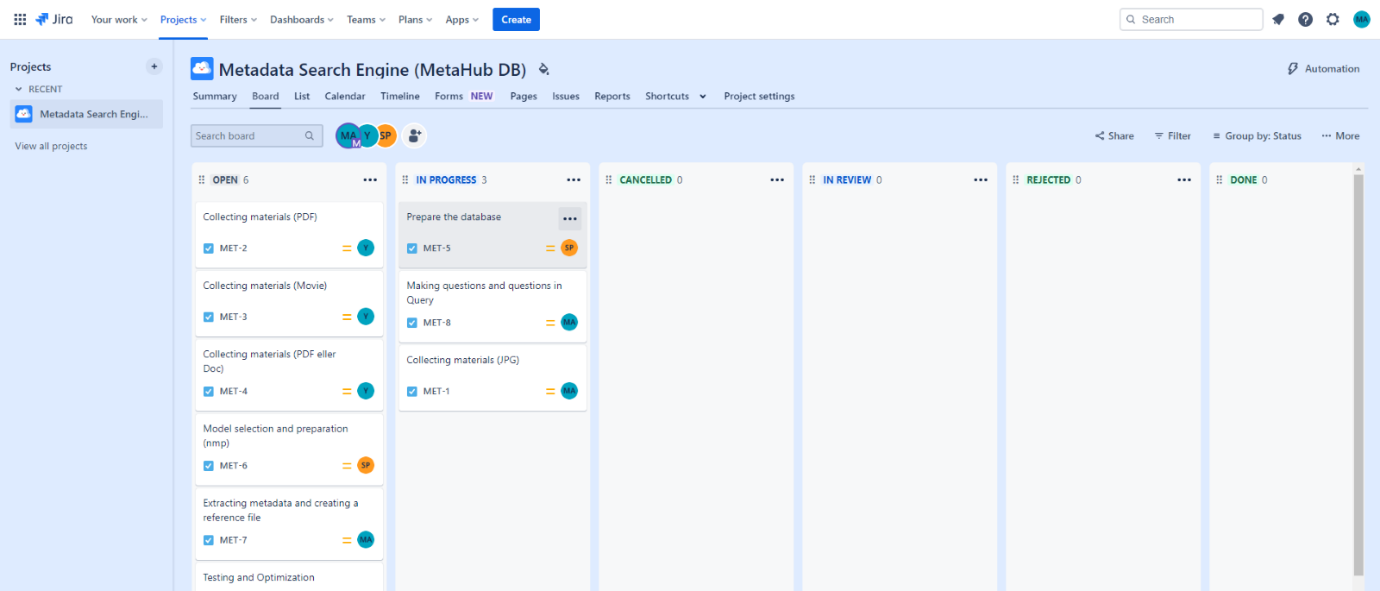
Communication:

We established clear communication channels using Teams, ensuring effective information exchange and prompt issue resolution.



Project Management:

Jira (Kanban) was employed for task organization, progress tracking, and task assignment. This tool helped us maintain clarity on project timelines and individual responsibilities.



Code Collaboration:

We adopted a feature branch strategy using Git. Each team member worked on a specific feature in a separate branch before merging into the main branch. Code reviews were regularly conducted to ensure code quality.

Meetings:

Regular team meetings were scheduled to discuss progress, challenges, and coordinate tasks. These meetings fostered a collaborative environment and allowed for quick problem resolution.

Project Management and Individual Contributions:

During the project, effective project management was crucial for coordinating tasks, tracking progress, and assigning responsibilities. We employed Jira(Kanban) to streamline these processes and ensure a well-organized development workflow. Each team member had specific roles and responsibilities within the project:

* **MohammadReza Abedi**: As the project manager, I took the lead in mapping out the project. This involved overseeing the overall project structure, coordinating task assignments, and ensuring alignment with project timelines.
* **Yinghua Zhu:** Ingwa played a crucial role in sourcing and collecting relevant materials. Their contribution was instrumental in gathering the necessary information for the project's implementation.
* **Shiva Paknegar:** Shiva took charge of collecting coding patterns essential for our development tasks. Their role was pivotal in establishing a foundation for our coding practices.
* **Meetings and Collaboration:** Regular team meetings provided a platform for collective decision-making. During these sessions, we drew conclusions on how to integrate resources and connect them to the established coding patterns. Collective participation and testing on models allowed us to refine our approach.

Throughout the process, Reza took on the responsibility of ensuring that the collected sources aligned seamlessly with the designated implementation patterns. This quality check was crucial for maintaining consistency in our work.

The collaborative efforts of the team members in their respective roles, combined with the use of Jira(Kanban), contributed significantly to maintaining clarity on project timelines and individual responsibilities.

Challenges and Solutions:

Challenges Encountered:

[List any challenges or obstacles faced during the project.]

Solutions Implemented:

[Describe how the team addressed the challenges and the solutions implemented.]

Testing and Optimization:

Thorough testing was conducted to ensure the search engine's accuracy and functionality. We implemented both unit testing and integration testing to identify and address any issues promptly. Additionally, optimization strategies were employed to enhance the search engine's performance as the database grew.

Documentation:

Comprehensive documentation was created, covering the codebase, database schema, and other relevant information. A README.md file was prepared to guide users on setting up and running the project.

Conclusion:

In conclusion, our collaborative effort resulted in the successful development of a metadata search engine that meets the specified requirements. The team's commitment, effective communication, and individual contributions collectively contributed to the project's success.

Future Recommendations:

For future projects of this nature, we recommend [Specify any recommendations or lessons learned during the project].

Report on Image Metadata Extraction and Database Insertion

**Technologies Used**

Programming Language: JavaScript (Node.js)

**Libraries:**

* exifr: Utilized for extracting metadata from images.
* fs: Employed for reading files from the file system.
* mysql2: Chosen as the MySQL client for Node.js.
* Database: MySQL

**Code Overview**

* **Metadata Extraction**

The script begins by importing necessary libraries and establishing a connection pool to the MySQL database. It utilizes the fs module to traverse through a directory containing files of various formats. For each file, it employs format-specific libraries or methods to extract relevant metadata.

* **Database Insertion**

Following metadata extraction, the script connects to the MySQL database using the established connection pool. It inserts the filename and associated metadata into a designated table. To maintain flexibility for different file formats, metadata is stored in a format-agnostic manner, often as a JSON string.

* **Error Handling**

The script incorporates robust error handling to manage potential issues during metadata extraction or database insertion. Errors are logged for later analysis and troubleshooting.

The extracted metadata provides information about the properties and characteristics of the images. Here's a summary of the key information extracted from the images:

Image: Image/107.jpg

Make: LGE

Model: Nexus 5X

Orientation: Horizontal (normal)

Resolution: 4032x3024

Exposure Time: 0.00131316 seconds

Aperture: f/2

ISO: 60

Focal Length: 4.67 mm

GPS Coordinates: Latitude 56.20307498149911, Longitude -4.719141858102828