DSCI-560 Assignment No. 5 - Part 2

Team Formation

Team name: HealthEdBot

Names & USC ID

Cici Chang (8706354957)

Daoyang Li (7168949829)

Yifan Yang(8386626967)

https://github.com/DSCI560

Minutes of the offline meeting on February 26

Participants: project team members

Conference topic:

We first discuss the importance of accurately extracting data from PDF for analysis, given the complexity of the PDF format. The problem is not simple, and various open source tools for PDF text extraction, such as Apache Tika and PDFMiner, are explored and their API and Python compatibility are discussed for easy integration. The importance of clean and consistent data sets for effective visualization is emphasized and preliminary data cleaning steps are decided upon. Evaluate server setups for different Linux distributions, emphasizing stability and community support. Ubuntu and CentOS are considered top candidates. Our previous experiments established the use of Ubuntu. The need for responsive web design is identified to ensure accessibility across various devices, HTML5, CSS3 and JavaScript frameworks such as Bootstrap are discussed.

Cici Chang to investigate PDF text extraction tools and begin extracting data from the provided PDFs.

Daoyang Li to set up the Apache web server and prepare the basic webpage layout.

Yifan Yang to research on MySQL database setup and data organization.

Minutes of the online meeting on February 27

Participants: project team members

Conference topic:

Cici provided an overview of the extracted data's structure, discussing the challenges faced with non-uniform data and potential solutions for normalization. Daoyang showcased the webpage's wireframe, emphasizing user experience and navigation. Discussed incorporating AJAX for dynamic data loading without page refresh. Yifan proposed a normalized database schema to handle well information efficiently, addressing the need for scalability and potential future data integration. Debated the merits of different SQL vs. NoSQL databases for this project, considering the structured nature of the data and the requirement for complex queries.

Cici Chang to finalize the data extraction and start preprocessing.

Daoyang Li to integrate OpenLayers into the webpage and plot base maps.

Yifan Yang to finalize the database schema and start importing the extracted data.

Minutes of the offline meeting on February 28

Participants: project team members

Conference topic:

This day's work is complex, and we evaluate progress in data preprocessing, focusing on the complexity of spatial data and its alignment with well coordinates. We spent a lot of effort carefully studying the different basemaps available in OpenLayers and their suitability for representing geological data such as terrain and satellite imagery. The application of geospatial data indexes for efficient querying and rendering of well locations is discussed. Explore integrations with third-party APIs to enrich well data with additional geological and environmental information.

Cici Chang to work on cleaning the data and ensuring consistency for web integration.

Daoyang Li to implement push pins on the map and popup mechanism for detailed information.

Yifan Yang to assist Daoyang with data integration and test database queries for the web application.

Minutes of the online meeting on February 29

Participants: project team members

Conference topic:

Reviewed the testing outcomes, identifying discrepancies in geolocation data which could be attributed to coordinate reference system mismatches. Discussed the common challenges in web mapping applications, such as handling large datasets and optimizing map rendering performance. Addressed the user interface considerations for the popup windows, ensuring that data is displayed in a user-friendly manner.

Explored the usage of web analytics tools to understand user interaction with the map features for future improvements.

Cici Chang to verify and correct any data-related issues that affect mapping features.

Daoyang Li to troubleshoot and refine the push pin and popup features on the map.

Yifan Yang to support testing and ensure the database provides accurate data for the web application.

Minutes of the online meeting on March 1

Participants: project team members

Conference topic:

Conducted a comprehensive review of the web application, ensuring all components functioned as intended and met the project's requirements.

Planned the structure of the demonstration video, focusing on clearly communicating the project's goals, methodologies, and outcomes.

Outlined a narrative for the video to guide viewers through the data journey, from extraction to visualization.

Discussed potential questions and feedback from viewers, preparing the team to respond effectively.

Cici Chang to prepare the final presentation and lead the video recording.

Daoyang Li to finalize the web application and ensure it is ready for the

demonstration.

Yifan Yang to compile all documentation and meeting notes for submission.

Finally we recorded a video to show our work.

Team name: HealthEdBot
Cici Chang (8706354957)
Daoyang Li (7168949829)
Yifan Yang(8386626967)