

Meeting Minutes - Lab 6: Text Extraction, Web Scraping, and Data Processing

Group: 6

February 18 - February 21, 2025

February 18, 2025

- Discussed project scope and tasks.
- Identified key technologies: Python, BeautifulSoup, Pandas, MySQL.
- Divided tasks among team members.
- Started with PDF text extraction script.
- Tested initial text extraction from sample PDFs.

February 19, 2025

- Improved text extraction logic to handle multiple PDFs.
- Began preprocessing extracted data (removal of missing values and formatting).
- Started developing the web scraping script for additional data collection.
- Encountered first connection issues with API requests.
- Debugged API request failures and adjusted parameters.

February 20, 2025

- Refined web scraping logic and handled edge cases.
- Completed CSV file generation with cleaned extracted data.
- Created MySQL database `Lab6_database` and `oilwell_data` table.
- Inserted test records into MySQL database.
- Validated data structure using phpMyAdmin.

February 21, 2025

- Fixed final issues with failed API calls.
- Reviewed and verified data stored in the database.
- Documented project workflow and finalized README file.

February 22, 2025

- Submitted project deliverables and reviewed final documentation.