Mid-Term Project Report

Project Title: Efficient XML Parsing and Database Integration Project

Team Members:

Sai Sumanth Boddu (Y00859727)

Manju Madala (Y00857071)

Faculty Coach: Dr. Feng George Yu

Project Description:

Our project aims to create a streamlined system for parsing XML data from the National Science Foundation (NSF) award data repository and integrating it into a MySQL database. The NSF award data repository contains valuable information about research grants and awards funded by the NSF across various domains of science and engineering. By parsing this data and storing it in a structured database format, we aim to provide users with easy access to this wealth of information for research, analysis, and decision-making purposes.

Data Source:

The data source for our project is the NSF award data repository, accessible via the following link: NSF Award Data Repository. This repository provides XML files containing detailed information about NSF awards, including award numbers, project titles, principal investigators, funding amounts, project abstracts, and more.

XML Schema:

To ensure consistency and accuracy in parsing the XML data, we will utilize the XML schema provided by NSF for their award data. The XML schema defines the structure and elements of the XML files, ensuring that the data is parsed correctly and can be seamlessly integrated into our database. The XML schema can be accessed via the following link: NSF Award Data XML Schema.

Milestones Completed:

Research and Familiarization: We successfully conducted research on XML parsing libraries and database integration techniques, familiarizing ourselves with Python's xml.etree.ElementTree library and MySQL/MariaDB.

Design and Development of XML Parsing Module and Database Schema: We designed and developed the XML parsing module using Python, ensuring compliance with the provided XML schema. Additionally, we defined the database schema and created tables dynamically based on the extracted information.

Backend Logic Implementation: Backend logic for file uploads, data parsing, and database insertion has been implemented using Node.js and Express.js. We have successfully integrated the parsing module with the database, allowing for seamless insertion of parsed data.

Milestones to Complete:

Frontend Development: Develop a user-friendly web interface using HTML, CSS, and JavaScript for facilitating user interaction with the system.

Testing and Performance Optimization: Conduct rigorous testing, including unit testing, integration testing, and performance testing, to identify and address any potential issues or bottlenecks. Implement performance optimization techniques to enhance system scalability and responsiveness.

Documentation and Project Submission: Finalize documentation, including user manuals and technical specifications, and prepare for project submission.

Initial Results:

The XML parsing module has been successfully developed and tested, demonstrating efficient extraction of relevant information from NSF award data XML files.

Backend logic for file uploads, data parsing, and database insertion has been implemented and integrated, allowing for seamless processing and storage of parsed data into the MySQL database.

Problems/Issues:

Frontend Development Delay: Due to unforeseen complexities in designing the user interface, the frontend development phase has been delayed slightly. We are actively working to address these challenges and expedite the development process.

Performance Optimization Challenges: While the system is functional, we have encountered some performance issues during initial testing, particularly with handling large volumes of XML files. We are exploring various optimization techniques to improve system performance and scalability.

Next Steps:

Prioritize frontend development to ensure timely completion of the user interface. Address performance optimization challenges through iterative testing and refinement. Finalize documentation and prepare for project submission.

Conclusion:

The project has made significant progress in achieving its objectives of efficiently parsing XML data from the NSF award data repository and integrating it into a MySQL database. Despite encountering some challenges, we remain confident in our ability to deliver a robust and user-friendly system by the project deadline. With continued dedication and collaboration, we are poised to successfully complete the remaining milestones and achieve our project goals.