Project Proposal.

ISO 15926-4 Interactive Visualisation Project

CLIENT: Professor Melinda Hodkiewicz

TEAM: Team - 18

1. AIM & BACKGROUND

Project Aim

The aim of this project is to develop an interactive visualization tool for the equipment reference data library described in the ISO 15926-4 Standard. This tool will empower members of the ISO TC 184 SC4 committee, including Prof. Melinda Hodkiewicz, to efficiently interrogate and visualize hierarchical data relationships through a web-based interface. The tool is slated for demonstration at the ISO meeting in Stavanger, Norway, in late October.

Background

Currently, the committee members face challenges in manually searching through extensive data records to locate relevant information, leading to inefficiencies and time-consuming processes. The data is available at the SPARQL endpoint (https://data.15926.org/sparql), but the lack of an intuitive interface makes navigation through these large datasets difficult.

The proposed solution is to create a web-based interface that integrates seamlessly with the existing SPARQL database, allowing users to query the database and visualize the hierarchical relationships inherent in the equipment reference data.

2. KEY BENEFITS

- **Efficiency:** The tool will drastically reduce the time and effort required to search and visualize data, significantly improving productivity for the committee members.

- **Intuitive Interface:** The interactive visualization tool will offer a user-friendly interface that simplifies navigation through complex data relationships, making it accessible even to non-technical users.

- **Enhanced Decision-Making:** By providing clear and accessible visualizations of hierarchical data, the tool will support better decision-making processes.

- **Demonstration at ISO Meeting:** The tool will be showcased at the ISO meeting in Stavanger, Norway, demonstrating its practical application and benefits to a broader audience.

3. DELIVERABLES

Interactive Visualization Tool

- **Overview:** A web-based interface built using Streamlit that allows users to query and visualize hierarchical data relationships from the ISO 15926-4 equipment reference data library.

- **Search and Visualization Features:**

  - **Search by Unique Name:** A search bar enabling users to input a unique name and find specific nodes within the hierarchy.

  - **Parent Node Display:** An option allowing users to specify the number of parent nodes they wish to view for a selected unique name node.

  - **Child Node Display:** An option for users to specify the number of child nodes they want to display for a selected unique name node.

  - **Node Details Display:** When a node is clicked, detailed information about that node will be shown on one side of the screen, allowing for quick access and review of associated details.

  - **Interactive Exploration:** Users can expand and collapse branches of the hierarchy to explore parent and child nodes in greater detail.

These search bars will allow users to customize their view of the hierarchical data, enabling detailed exploration of the relationships between nodes. Both parent branches and child branches of the hierarchy can be expanded and collapsed for detailed viewing.

In addition, when a user clicks on a node within the visualization, detailed information about that node will be displayed on one side of the screen. This feature will allow users to quickly access and view relevant details associated with the selected node, enhancing the data exploration and analysis experience , and must be adaptable to different screen sizes.

Deployment on AWS

- **Overview:** The visualization tool will be deployed on Amazon AWS to ensure it is easily accessible and scalable for broader use.

User Documentation

- **Content:** Comprehensive documentation will be provided to guide users in effectively utilizing the visualization tool. This will include instructions on search functionality, data filtering, and interpreting visualized relationships.

4. PROPOSED METHODS:

TASK 1: Enhanced Visualization and Exploration of Class Datasets

To address the challenges faced by the committee in navigating and interpreting the complex data, we will employ several key methods and technologies. First, to prepare and integrate data, we will establish a connection to the SPARQL endpoint at `https://data.15926.org/sparql` to enable real-time querying and data retrieval and use SPARQL queries to extract relevant data, focusing on hierarchical relationships as outlined in the ISO 15926-4 Standard. Second, to parse and organize data, we will parse the retrieved SPARQL query results to structure the data into a format suitable for visualization that is a CSV file and then organize the data into a hierarchical format to accurately represent parent-child relationships within the dataset.

TASK 2: Interactive Web-based Interface Development

In developing the interactive web-based interface, we will utilize Streamlit for rapid development and deployment of the web application, taking advantage of its simplicity and interactive capabilities. The user interface will be designed to be intuitive and user-friendly, allowing users to interact with the hierarchical data easily. This will involve implementing search bars for querying nodes by unique names and providing options for specifying number of parent and child nodes to view.

TASK 3: Data Visualization

For data visualization, we will leverage visualization libraries such as Networkx or D3.js to enable dynamic and interactive visualizations within the Streamlit app. Hierarchical visualization techniques, such as tree diagrams or sunburst charts, will be used to effectively display data relationships. The visualization will feature interactive elements, including expandable and collapsible branches, allowing users to explore the hierarchy interactively. Additionally, clickable nodes will display detailed information about the node, facilitating in-depth analysis.

TASK 4: Testing and Iteration

We will conduct user testing sessions with members of the ISO TC 184 SC4 committee to gather feedback on the tool's usability and functionality. Based on the feedback, we will iterate on the interface design and features to ensure that the tool meets the needs of both technical and non-technical users. Performance optimization will be a key focus, as we aim to ensure that SPARQL queries and data retrieval are efficient, resulting in fast loading times and smooth interactions within the application.

TASK 5: Deployment and Demonstration

The final version of the interactive visualization tool will be deployed on a cloud-based platform to ensure accessibility and scalability. In preparation for the ISO meeting in Stavanger, Norway, we will prepare a demonstration of the tool, highlighting its practical application and benefits to the committee members and a broader audience.

5. MVP AGREEMENT

Minimal Viable Product (MVP)

- **Agreed Features:**

    - Basic search functionality for searching a particular node using the "Unique Name", and two more search bars allowing users to specify the number of parent nodes and child nodes they wish to view for a selected unique name node.

    - Visualisation of hierarchical data with parent and child node display options.

    - Node information to be displayed on clicking it.

    - We cover the "Functional Object" endpoint dataset.

    - The webpage must be adaptable to different screen sizes.

- **Client Agreement:** The MVP has been reviewed and agreed upon with the client, ensuring alignment with the project's goals and timeline. [Hyperlink to the MVP agreement meeting notes]

6. DELIVERABLE TIMELINE & APPROXIMATE TIME COST

Week 1-3: Meet with the client and ensure the requirements.

Week 4: Establish a connection to the SPARQL endpoint, extract relevant data, and convert it into a structured format for visualization.

Week 5: Organize the parsed data into a hierarchical format to accurately represent parent-child relationships and ensure data integrity.

Week 6: Develop the web application interface using Streamlit, implementing search functionality and integrating data visualization components.

Week 7: Create interactive visualizations using libraries like Networkx or D3.js incorporating expandable branches and clickable nodes for detailed exploration.

Week 8: Conduct user testing sessions to gather feedback, iterating on the design and features to enhance usability and functionality.

Week 9: Optimize performance by refining SPARQL queries and interface interactions, ensuring fast loading times and a smooth user experience.

Week 10: Deploy the tool on a cloud-based platform and prepare for the demonstration at the ISO meeting in Stavanger, Norway.

7. USER EXPERIENCE (UX) ANALYSIS

Target Users

- **Primary Users:** Members of the ISO TC 184 SC4 committee, including Prof. Melinda Hodkiewicz.

UX Considerations

- **Ease of Use:** The interface is designed to be user-friendly, with clear navigation and easily accessible functionalities.

- **Responsive Design:** The tool will be optimized for different devices, ensuring usability across various screen sizes and resolutions.

- **Accessibility:** Ensure the features are accessible via keyboard shortcuts to accommodate users with different needs.

- **Intuitive Data Interaction:** Enable users to interact directly with visual elements, such as clicking on a node in a hierarchical graph to explore its details or relationships.

- **Future Plans:** Additional UX enhancements will be explored in later stages of the project based on user feedback from the initial prototype.

8. TECHNOLOGY CHOICES

Selected Technologies

- **Streamlit:** Chosen for its simplicity in creating web applications with Python.

- **SPARQL:** Utilized for querying the ISO 15926-4 database.

- **Amazon AWS:** Selected for deploying the application due to its reliability and scalability.

Rationale

- **Group Skills:** The technologies chosen align with the team's existing skills in Python and web development.
- **Project Risks:** AWS offers a secure and scalable environment, mitigating risks related to hosting and accessibility.

9. PRELIMINARY SECURITY THREAT MODELING

STRIDE Analysis

- **Threats Identified:** Preliminary STRIDE analysis has identified potential risks, including spoofing, tampering and information disclosure.
- **Mitigation Strategies:** Implementing secure authentication mechanisms and data encryption to address identified threats.
- **Next Steps:** The security analysis will be reviewed and updated throughout the project lifecycle

10. PROJECT ARTIFACTS AND HYPERLINKS

- **GitHub:** Link to Repository
- **MS Teams:** Link to Project Files and Discussions
- **Notion:** Link to Notion Board

11. CLIENT REVIEW AND FEEDBACK

Client Review Section

- **Feedback Placeholder:** The client's feedback will be documented here post-submission.

Client Communication

- **Approval:** Document the client's approval of the submission and any additional comments.

12. GROUP MEMBERS AND STUDENT NUMBERS

Team Members and Responsibilities

- **Team Member 1: Xudong Ying (21938264)**

  - **Tasks:** Documentation, Data Retrieval, Back-End.

- **Team Member 2: Manish Varada Reddy (23817492)**

  - **Tasks:** StreamLit, Proposal Report, Front-End.

- **Team Member 3: Yu Xia (24125299)**

  - **Tasks:** Design a logic and code for "Data Retrieval" from the end-point, Back-End.

- **Team Member 4: Zihan Zhang (23956788)**

  - **Tasks:** D3.js, Back-End.

- **Team Member 5: Melo Xue (23955182)**

  - **Tasks:** Proposal Report, Front-End.

- **Team Member 6: Shanmugapriya Sankarraj (23872782)**

  - **Tasks:** Proposal Report, Front-End.