SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name: **Masters of Science (Data Science)**

Subject Name: **Project 1 (MCST1043)**

Student Name: Yang Mu

Metric Number: MCS241045

Student Email & Phone: yangmu@graduate.utm.my number:1162302346

Project Title: Construction and visualization of metadata framework of COVID-19 scientific dataset- Re3data.org

Supervisor 1:

Supervisor 2 / Industry Advisor(if any):

SECTION B: Project Proposal

Introduction:

The COVID-19 pandemic, which began in December 2019, has had a profound impact on the world, causing over 500 million infections and over 6.22 million deaths. This global health crisis has spurred an unprecedented surge in scientific research, generating a vast amount of data related to the virus, its transmission, treatment, and societal impact. However, effectively managing and utilizing this wealth of information presents a significant challenge.

Problem Background:

While the availability of extensive COVID-19 research data is a valuable resource for understanding the pandemic and developing solutions, it also poses significant challenges. The sheer volume of data, often scattered across various sources, makes it difficult to discover, access, and integrate relevant information. Moreover, data duplication, redundancy, and inconsistent metadata standards hinder efficient data analysis and knowledge discovery.

Problem Statement:

The challenge lies in developing effective strategies for organizing, managing, and utilizing the vast amount of COVID-19 research data. This includes addressing issues of data redundancy, inconsistency, and lack of discoverability, ultimately enabling Researchers use this data for more comprehensive and impactful studies, as well as for the prediction and prevention of major infectious diseases.

Aim of the Project:

This research aims to address this challenge by constructing a robust metadata will framework specifically designed for COVID-19 scientific datasets. This framework will provide a standardized structure for describing and organizing data, enhancing its discoverability and facilitating its reuse for further research. Furthermore, the project explore the use of knowledge graphs to visualize this metadata, creating a more intuitive and interactive representation of the relationships between data elements.

Objectives of the Project:

Collect and organize metadata from COVID-19 scientific datasets.

Construct a metadata framework for COVID-19 scientific datasets.

Utilize Protégé software to build an ontology for scientific datasets.

Store the constructed knowledge graph using the Neo4j graph database.

Enable the display and reuse of COVID-19 scientific dataset metadata.

Scopes of the Project:

This research focuses on building a metadata framework and conducting visualization studies using COVID-19 scientific datasets from Re3data.org as an example. The project will explore the application of this framework and visualization techniques to a specific set of COVID-19 scientific datasets, demonstrating its potential for enhancing data management and research within this critical domain.

Expected Contribution of the Project:

Methodological Advancement: This project will contribute to the development of the robust methodologies for organizing and managing scientific datasets, particularly in context of COVID-19 research.

Knowledge Organization: The constructed metadata framework and knowledge graph will provide a valuable resource for researchers seeking to understand the complex relationships between COVID-19 datasets and their underlying knowledge.

Research Enhancement: By facilitating the discoverability, accessibility, and reuse of COVID-19 scientific datasets, this project will empower researchers to conduct more comprehensive and impactful studies, ultimately contributing to a better understanding of the pandemic and the development of effective solutions.

Project Requirements:

Software: Neo4j, Protégé

Hardware: Computer, network, data

Technology/Technique/ Methodology/Algorithm: Metadata framework construction, knowledge graph construction, ontology building, data visualization

- Collect and organize metadata from COVID-19 scientific datasets.
- Construct a metadata framework for COVID-19 scientific datasets,encompassing external features, content features, and sharing features.
- Utilize Protégé software to build an ontology for scientific datasets, defining core concepts and attribute relationships.
- Store the constructed knowledge graph using the Neo4j graph database, enabling query retrieval and reasoning for entities and their relationships.

Type of Project (Focusing on Data Science):

[ ] Data Preparation and Modeling

[ ] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[✓] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

Status of Project:

[✓] New

[ ] Continued

If continued, what is the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

[✓] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: Yang Mu

Yang Mu | 07/04/2025

Signature | **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1: .....

Signature | **Date**

Name of Supervisor 2 (if any): .....

Signature | **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL
[ ] CONDITIONAL APPROVAL (Major)*
[ ] CONDITIONAL APPROVAL (Minor)
[ ] FAIL*

***** Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1: .....

Signature ..... **Date** .....

Name of Evaluator 2: .....

Signature ..... **Date** .....