Suggested Report Structure
MOD004979

Assessment Title:
Semantic Modelling for IMDB Movie Dataset (example)

Team:
Name and (SID)
John Doe (123456)
Bob Alex (767687)

Table of Contents

1. Introduction

A brief introduction for the overall project. It is not graded but will add good starting point for the report.

2. Concept and Requirement Analysis (10%)

This task is a **GROUP** work. You provide here the aim and rationale for your choice of application considering if the application area is a suitable one for a semantic-based application. Provide list of questions which maybe asked from this system (at-least six questions per team member e.g. for a team of 4, you will have at-least 24 different questions).

2.1 Aim of the Project

2.2 Gap in the Knowledge (Rationale)

2.3 List of Questions

No.	Questions
1	Name all movies released after 2017

Table 1: List of Questions

3. Design (Semantic Model for Movies Dataset) (25%)

This task is a **GROUP** work. Use Protégé to create an Ontology for your dataset. The Ontology should have good range of classes (minimum 20 main classes), properties (minimum 25) and individuals. Explain the constraints and axioms and your process of designing the Ontology Model.

3.1 Ontology Summary

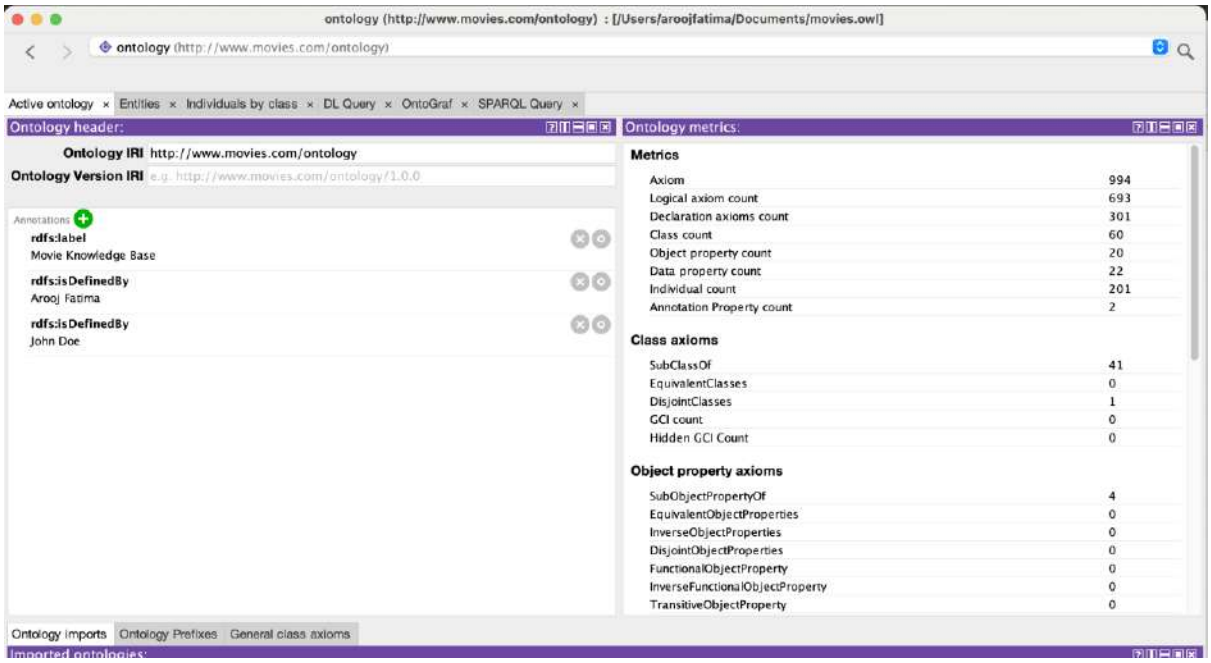

Figure 1: Movie Dataset Ontology Overview

3.2 Classes

Describe your classes. Expand hierarchies and provide screenshots.

3.3 Properties

Provide list of your data and object properties with explanation and screenshots from Protégé.

3.4 Individuals

Provide details of individuals created, excel sheets' snapshots and any rules written to import data. Provide detail of any other method (if used) to import data).

3.5 Axioms

List all axioms with examples from Protégé.

Provide as many screenshot as you can since marks will be based upon the reports only.

4. Requirements Mapping (05%)

This task is a **GROUP** work. Map all the questions created in first task 'Concept and Requirement analysis' to the ontology model. provide appropriate explanation. You may create a table to demonstrate this mapping.

4.1 Mapping Table

Question	Mapping
Name all movies released after 2017	This question can be answered using Movie class and data properties (name, date_of_release).

Table 2: Requirement Mapping

5. Evaluation and Use (20%)

This is an **INDIVIDUAL** task. The marks for this section will be given for the SPARQL SELECT queries. You should choose five questions from the list of questions created in Step 1. Be considerate while choosing the questions to ensure the questions let you demonstrate your grasp on SPARQL. Please provide queries with the screenshots for the results (either run on Protégé or your SPARQL end-point). Provide a brief explanation for each query. You should not attempt more than five questions. Any extra queries/questions in this section will be ignored. Any similar pattern queries will be considered as same query. Choose queries which are different from each other and shows range of SPARQL features.

Provide as many screenshot as you can since marks will be based upon the reports only.

5.1 Use case 1

Question:
Name all movies released after 2017

SPARQL Code and Output:

Figure 1: SPARQL query and result

Explanation:
The query list names and release year of all the movies in the dataset. The query will give better results if we use FILTER for 2017 ☺

5.2 Use case 2

5.3 Use case 3

5.4 Use case 4

5.5 Use case 5

6. Implementation (30%)

This is **INDIVIDUAL** task. Write code for a user interface (using Python) which takes Natural Language queries (text), converts user query to a SPARQL query which runs on your dataset and displays results. Marks for this section will be given based upon the Natural Language Processing Techniques, connection to the dataset, results display, critical exploration/testing and quality of code. The code should be dynamic and the examples in the report should include at-least 5 SPARQL queries.

Higher marks will be awarded for good literature references and testing various queries and NLP techniques.

Provide as many screenshot as you can since marks will be based upon the reports only.

6.1 Connection to the Dataset:

Provide evidence that your code is connected to your dataset. You may utilise screenshots from Fuseki and Jupyter Notebook.

6.2 NLP Techniques:

List and explain the use of NLP techniques in your project.

6.3 Overall Code:

Explain the flow of your code to make reader understand the concepts.

6.4 Code Validation:

Provide a step by step demo / flow (using screenshots) for your 5 NL queries to show how your code process the query and produce output.

7. Conclusion and Future Work

This step again is not graded but it provides a closure to the project. I wish you all the very best for the module and your overall learning journey.