Knowledge Graphs with Large Language Models

Assignment 1: Knowledge Graph Schema Design

Introduction

In this assignment you are asked to consider the small movie knowledge graph we saw in class, and expand it so that it can answer the following competency questions:

- Which movies have been produced in the United States?
- Which movies are comedies and which are romantic comedies?
- What rating did a reviewer give to a given movie and what was their review text?

Task 1 (50%)

Expand the Neo4j version of the knowledge graph (available in the file "movies_graph_neo4j.dump") so that it can answer the above competency questions. For that, you can use the Neo4J AuraDB.

Verify that the questions can be answered by expressing the competency questions in Cypher and running them against sample data. Describe and explain the modeling decisions you had to make.

Task 2 (50%)

Expand the RDF(S)/OWL version of the knowledge graph that is in file "movies_graph_rdf_owl.owl", so that it can answer the competency questions. You can use Protege or any other tool you want.

Verify that the questions can be answered by expressing the competency questions in SPARQL and running them against sample data. Describe and explain the modeling decisions you had to make.

Guidelines and remarks:

- Consider the competency questions all together, not individually. That is because a design decision that is appropriate for one question might not be appropriate for another.
- Not only are you allowed but you are actively encouraged to use an LLM to help you with these tasks. However, if you do that, you need:
    - to be critical against the solutions that the LLM might give you and correct/adapt them as needed,
    - to include in your report the prompts and the prompting strategy you have followed.
- The schemas in tasks 1 and 2 need to be semantically correct but also adequately understandable to a human. That means that you should pay particular attention to the naming and description of their elements.
- Along with the schema I want to see the modeling thinking process you followed and the decisions you made, especially in cases where multiple modeling choices were available (e.g. deciding to model an entity characteristic as a relation instead of an attribute).

Deliverables

- A file with the expanded RDF/OWL version of the movies graph
- A dump file with the expanded Neo4j version of the movies graph, exported from AuraDB
- A report describing the process you followed, the design decisions you made, the LLM prompts you might have used for assistance, and the queries you used for testing.