CW1: Knowledge Engineering

|  |  |
| --- | --- |
| **Weight:** | 15% |
| **Stage 1 Deadline:** | 15:30 (GMT) Wednesday 2 February 2022 |
| **Submission** | Turtle OWL file through Canvas |
| **Stage 2 Deadline:** | 15:30 (GMT) Monday 21 February 2022 |
| **Submission:** | Report submitted through Canvas  Turtle OWL file through Canvas |

Table of Contents

[Overview 2](#_Toc90045458)

[Coursework Tasks 2](#_Toc90045459)

[Scenario 2](#_Toc90045460)

[Task 1: Ontology Development [30 Marks] 2](#_Toc90045461)

[Task 2: Querying your Ontology 2](#_Toc90045462)

[Query 1 [10 Marks] 3](#_Toc90045463)

[Query 2 [10 Marks] 3](#_Toc90045464)

[Query 3 [10 Marks] 3](#_Toc90045465)

[Task 3 (Level 11 Students only) [10 Marks] 3](#_Toc90045466)

[Stage 1: Individual Work 3](#_Toc90045467)

[Activity 3](#_Toc90045468)

[Collaboration and Plagiarism 3](#_Toc90045469)

[Stage 2: Group Work 4](#_Toc90045470)

[Activities 4](#_Toc90045471)

[Report Content 4](#_Toc90045472)

[Collaboration and Plagiarism 4](#_Toc90045473)

# Overview

The coursework involves developing an ontology to represent a specific domain according to the criteria outlined in this document. You are then required to query your ontology together with an external knowledge graph.

There are two stages to the coursework. If there is no submission for Stage 1, then a 30% penalty will be applied to the individual’s mark for the groupwork[[1]](#footnote-1).

# Coursework Tasks

## Scenario

WattTours offer tours to visit a series of cities. They have aspirations of offering tours to a large number of cities and countries, but do not have the resources to keep up-to-date information about all their destinations. They want to exploit the power of linked data through general purpose knowledge graphs to provide additional information about their tour destinations.

## Task 1: Ontology Development [30 Marks]

Develop a small ontology in Protégé to represent the needs of WattTours. You can reuse existing classes and properties to inherit from, but your ontology should be self-contained. The ontology will be assessed for following good data sharing practices, i.e. the assessment will cover metadata provided about the ontology as well as the concepts and annotations within it (submitted ontologies will be run through the [OOPS Ontology Evaluator](http://oops.linkeddata.es/)).  
*Hint: The* [*Time Ontology*](https://www.w3.org/TR/owl-time/) *defines classes for durations and time units.*

The ontology should cover:

* Tours which visit a number of cities and extend for a number of days, e.g. you may have a 3 day tour that visits 2 cities or a 5 day tour that visits 3 cities. Capture the name and description of the tour, and the cities visited.   
  *(Do not worry about the order that the cities appear in the tour, just capture which cities are included in the tour.)*
* Basic details of the cities and countries covered in their tours.
  + For cities this would include the name of the city and which country it is located in.
  + For countries this would include the name of the country and its capital city.
  + Include a link to the representation of the city or country in an external knowledge graph such as Wikidata.

Populate your ontology with one tour per group member. A tour should cover three or more cities. The tours may share cities, but at least two cities per group member should be provided.

## Task 2: Querying your Ontology

Use an existing knowledge graph, e.g. Wikidata, to supplement the information in your ontology. You should not store the content of the external knowledge graph in your ontology.

For each of the following information needs, write a single SPARQL query to retrieve the data from the external source(s). For each information need, provide the federated SPARQL query that combines data from your local ontology with the external knowledge graph and show the response that you receive from Wikidata.

*Hint: Note that Wikidata and DBpedia both use different identity IRIs to those that are displayed on the web pages. These IRIs use HTTP rather than HTTPS and are of the form:*

* *DBpedia: http://dbpedia.org/resource/*
* *Wikidata: http://www.wikidata.org/entity/*

*If you store the IRIs as a data property then you will need to convert these to IRIs in your SPARQL queries. To do this you will need the* [*IRI*](https://www.w3.org/TR/sparql11-query/#func-iri) *and* [*STR*](https://www.w3.org/TR/sparql11-query/#func-str) *functions.*

### Query 1 [10 Marks]

Provide the name for the cities in your ontology in two non-English languages.

*Hint: Labels can be retrieved from Wikidata using rdfs:label property.*

### Query 2 [10 Marks]

Populate an information box for your cities containing the following data *where it is available*:

* Size (area)
* Population size
* Time zone offsets (cities may have more than one)
* Official website
* Number of twinned cities

*Ensure that your query answers are intelligible to human users.*

### Query 3 [10 Marks]

Provide the number of art galleries, museums, and botanical gardens in the city.   
*Note that there are several different types of museums on Wikidata.*

## Task 3 (Level 11 Students only) [10 Marks]

Discuss the advantages and disadvantages to WattTours of relying on an external knowledge graph.   
*(To get full marks you should include literature references to support your claims.)*

# Stage 1: Individual Work

This part of the coursework is to be conducted individually.

## Activity

Develop your own version of the ontology within Protégé.

## Collaboration and Plagiarism

Submissions for Stage 1 must be your **own** work. If some text or code in the coursework has been taken from other sources, these sources must be properly referenced. Failure to reference work that has been obtained from other sources or to copy the words and/or code of another student is plagiarism and will be reported to the School's Discipline Committee. If an individual is found guilty of plagiarism, the penalty could involve voiding the course.

Students must never give hard or soft copies of their coursework reports or code to other students, except when working in Stage 2 of the coursework. Students must always refuse any request from another student for a copy of their report and/or code. Sharing a coursework report and/or code with another student is collusion and will be reported to the School's Discipline Committee. If found guilty of collusion, the penalty could involve voiding the course.

# Stage 2: Group Work

This part of the coursework is to be conducted in groups of three. Each group must consist only of Level 11 students. Groups must be registered through Canvas by the end of Week 3.

**All members must contribute to the group work** as experience gained in these topics will help your understanding of the topic (and in the examination). The purpose of working in a group is to allow you to discuss the concepts covered in the work and deepen your understanding.

## Activities

Within your group, share your submissions for Stage 1. Discuss the merits of each design and use these to inform an agreed ontology within your group. Implement your consolidated ontology in Protégé.

Collectively work on the remaining tasks.

## Report Content

**Please** **provide a summary in your report stating the contributions of each group member**. If necessary, marks will be adjusted if some students have not participated enough.

Your report must summarise the discussions that led to your consolidated design. You may find it useful to include snippets from your individual submissions to show how you came to a consolidated design.

You should also document your ontology and how it satisfies the requirements detailed above. Include details of any design decisions that you have made, e.g. whether to include terms from an existing ontology, and justify these.

For Task 2, you should provide the SPARQL query as text in your document (not a screenshot), and a screenshot of the results obtained (ensure that your screenshot is readable; screenshots that are too small, fuzzy, or the wrong orientation will not be awarded marks).

## Collaboration and Plagiarism

Coursework reports and code must be the group’s own work. If some text or code in the coursework has been taken from other sources, these sources must be properly referenced. Failure to reference work that has been obtained from other sources or to copy the words and/or code of another student is plagiarism[[2]](#footnote-2) and will be reported to the School's Discipline Committee. If a group is found guilty of plagiarism, the penalty could involve voiding the course.

Students must never give hard or soft copies of their coursework reports or code to students in another group. Students must always refuse any request from another student not in their group for a copy of their report and/or code. Sharing a coursework report and/or code with another group is collusion and will be reported to the School's Discipline Committee. If found guilty of collusion, the penalty could involve voiding the course.

1. If there are mitigating circumstances that prevent your engagement with Stage 1, then please inform the lecturer so that appropriate adjustments can be made. [↑](#footnote-ref-1)
2. Heriot-Watt guidelines on plagiarism <https://www.hw.ac.uk/students/studies/examinations/plagiarism.htm> [↑](#footnote-ref-2)