



IN3067/INM713 Semantic Web Technologies and Knowledge Graphs

Coursework Project (Part 1): Ontology Modelling (Task OWL)

Ernesto Jiménez-Ruiz

Academic course: 2023-2024

Updated: February 5, 2024

1 Introduction

DEADLINE: Sunday, 3 March 2024, 5:00 PM

This first part of the coursework involves the creation of an ontology that covers the knowledge of a given domain. **This assignment constitutes 20% of the total mark for the IN3067/INM713 module.**

A **knowledge scientist** is in charge of adding context to the data to make it more useful, accessible, clean, reliable and ready to be used by downstream analytics or AI-based systems. In this Part 1 of the coursework you will adopt the role of a knowledge scientist to create an ontology that formalises the domain of a given data set in tabular format (*e.g.*, CSV file).

1.1 Working in pairs or individually

Work in pairs will be allowed this year, so you can work individually or team-up with one of your peers. Given the experience from previous years, working in pair will enhance the learning process as the module brings together students with different skills and backgrounds.

You should let me know by **February 23** if you are planning to work individually or in pairs (together with the name of the members) for Part 1 of the coursework. **Use the form in moodle.** **Groups formed after the deadline may not be accepted.**

1.2 Extenuating circumstances

If you are not able to submit your coursework for any medical or personal reasons beyond your control, you should contact the course officers and fill an extenuating circumstances form. Please let me know if you are applying for EC.

1.3 Plagiarism and academic misconduct

We encourage group discussions beyond the formed pairs about the details and solutions of the lab sessions. The coursework, however, should be conducted by the formed pairs or individually. When submitting material in moodle a plagiarism check is performed checking submissions from previous, the current year and external resources.

Note that working in pairs means that you work “together” on all tasks. You can take the lead in some tasks, while your partner leads others; but both of you must have a good understanding of all tasks. The most important outcome is that you learn and be able to independently propose interesting (not necessarily perfect) solutions.

2 Coursework tasks

Note that each task (and subtask) has a defined weight (and subweight). For example, as Task OWL counts towards 20% of the overall coursework mark, Task OWL.A will count up to 14% (*i.e.*, the 70% of 20%, $0.2 \times 0.7 = 0.14$), while Task OWL.B 6% (with each of the subtasks OWL.B.x counting 0.6%) towards the overall module mark.

2.1 Ontology modelling in Protégé (Task OWL.A, 70%)

You will not only be marked with respect to the final produced output; but also according to the ontology modelling choices and the discussion around them. For example, the created ontology does not need to be perfect or exhaustive (*i.e.*, covering all pieces of information), but it should be representative.

2.1.1 Dataset

The dataset for the coursework is based on the kaggle dataset about *Pizza Restaurants and the Pizza They Sell*.¹ The dataset version to be used in Part 1 of the coursework **contains 500 data rows** and is available in Moodle. We will refer to this dataset as `cw_data`. Some tips:

- The data contains many things, so no need to be exhaustive. Some things can be ignored, and you may also decide to focus on some specific parts.
- The ontology needs to abstract the domain so in principle instances are not required. A few as an example/proof-of-concept may help; but it is not required to "manually" create an instance for each occurrence of a city, for example.
- In other cases there may be a modelling choice to add important elements as instances or subclasses. But at this abstraction stage and given problem, a hierarchy of classes is preferred.
- In the second part of the coursework is when you will need to (automatically) populate the given model ontology with the given tabular data; *e.g.*, instances of the class `City`.

2.1.2 Domain model task

Create a **small ontology**² that models the domain of `cw_data`. Use Protégé to develop the ontology. Save the ontology into `turtle` format (*e.g.*, `.ttl`) and `rdf/xml` format (*e.g.*, `.owl`).

The ontology should abstract knowledge about the domain, valid for the data at hand but potentially valid for other datasets (*e.g.*, tables).

Subtask OWL.A.1: Change namespace for the ontology to `http://www.city.ac.uk/inm713-in3067/2024/your_name/` (**5%**).

Subtask OWL.A.2: Create a prefix (*e.g.*, your initials `ejr:`) for the defined namespace (**5%**).

¹Kaggle dataset about pizza restaurants: <https://www.kaggle.com/datafiniti/pizza-restaurants-and-the-pizza-they-sell>. This kaggle dataset is a subset of a dataset provided by Datafiniti's Business Database.

²Note that the model solution for this coursework Part 1, and input for Part 2, will be much larger and comprehensive.

Subtask OWL.A.3: Reuse the annotation property `dc:creator` and annotate/indicate the ontology has been created by you (5%).

Subtask OWL.A.4: Create classes and organize them into a hierarchy (40%).

- Tip: between 15 and 25 classes.

Subtask OWL.A.5: Create properties and property restrictions as necessary (30%).

- Local scope: create existential and universal restrictions (*e.g.*, around 5).
- Global scope: domain and ranges (*e.g.*, around 5).
- Create appropriate property characteristics.

Subtask OWL.A.6: Create annotations (*e.g.*, labels or synonyms) for some entities (5%).

Subtask OWL.A.7: Create a new Restaurant concept (*e.g.*, a restaurant with the style of your country or city, *e.g.*, `PizzeriaDaErnesto`) and a new Pizza concept (*e.g.*, `PizzaDaErnesto`) with its expected ingredients (10%).

2.2 Ontology modelling on “Paper” (Task OWL.B, 30%)

Create meaningful OWL axiom for each of the following modelling tasks (10% each) using a text editor. Use DL or Manchester syntax to define the axioms. **Add and describe the created axioms in the report.** Use any domain (*e.g.*, anatomy, movies, etc.), but different from the Pizza, Food and Restaurants domain.

Subtask OWL.B.1 An axiom with an atomic subsumption.

Subtask OWL.B.2 An axiom with an universal restriction.

Subtask OWL.B.3 An axiom with an existential restriction.

Subtask OWL.B.4 An axiom with a union.

Subtask OWL.B.5 An axiom with an intersection.

Subtask OWL.B.6 An equivalence axioms with a cardinality restriction.

Subtask OWL.B.7 An axiom stating two concepts cannot have a common instance.

Subtask OWL.B.8 A property chain axiom.

Subtask OWL.B.9 A role assertion axiom and a valid inverse.

Subtask OWL.B.10 A class assertion axiom where the class is complex.

Subtask OWL.B.11 (optional) A combination of axioms that makes the ontology to be outside OWL 2.

3 Submission guidelines

You need to submit all the material in Moodle, in the dedicated “Submission Area” (*i.e.*, *Coursework Assignment (Part 1) - 20%*). **Please follow the instructions carefully, otherwise the submission may not be marked.**

1. **Ontology** in both **ttl** (.ttl) and **rdf/xml** (.owl) formats.
2. **Video** where you describe the ontology and modelling choices.³ You can use slides as support and/or describe the ontology directly on the Protégé interface. The video should not be longer than 5 minutes (strict limit) and focus on the key modelling choices. In **mp4 format** or any other format that can be reproduced by a Web browser.⁴ In the unlikely case the video does not fit in moodle (>200Mb), please upload them in OneDrive (by the deadline) and submit a text file with the link to the video.
3. **Report**. A short document (maximum 3 pages) in **PDF** format, focusing of the main modelling choices in Task OWL.A, and containing the axioms of Task OWL.B.
4. The **ontology text code** in ttl and rdf/xml formats into a **single PDF file**. This is a requirement for auditing purposes and to run TurnItIn (to identify plagiarism) on the ontology code. Please make sure the text in the generated PDF is “selectable” (not an image).

3.1 Individual work

Please submit the files individually, **not in a zip file**, and provide short but **meaningful names**. That is, you will need to **submit 5 files** in total, for example:

- *Ernesto_Ontology.ttl*,
- *Ernesto_Ontology.owl*,
- *Ernesto_Video.mp4* (or *Ernesto_linkTo_Video.txt*),
- *Ernesto_report.pdf*, and
- *Ernesto_OntologyCode.pdf*.

3.2 Work in pairs

You need to create two independent videos where each member describes the ontology and key modelling choices.

Both members **MUST** submit material in moodle as follows:

³Two videos in the case you are working in pairs.

⁴To record the video, you can start a Zoom or Teams call with just yourself, share the screen, and record the meeting. I can help on this if necessary.

- **Member 1:** Please submit the files individually, **not in a zip file**, and provide short but **meaningful names**. That is, you will need to **submit 5 files** in total, for example:

- *GroupName_Ontology.ttl*,
- *GroupName_Ontology.owl*,
- *Member1Name_Video.mp4* (or *Member1Name_linkTo_Video.txt*),
- *GroupName_report.pdf*, and
- *GroupName_OntologyCode.pdf*.

- **Member 2:** Please only **submit your video file**; e.g., *Member2Name_Video.mp4* (or *Member2Name_linkTo_Video.txt*),

Each group decides internally who acts as Member 1 or Member 2.