

This question paper consists
of 5 printed pages each
of which is identified by the Code
Number COMP5860M01

*This is an open book examination.
Any written or printed material is permitted.*

© UNIVERSITY OF LEEDS

School of Computing

May 2018

COMP5860M01

Semantic Technologies and Applications

Answer all 3 questions

Time allowed: 2 hours

Question 1

A government organisation is reviewing its training content management strategy. The analysis of the current state has identified that there were numerous cases of duplications and inconsistencies, limited content reuse, poor effectiveness and efficiency in content finding by users. To mitigate these problems, the organisation invests in a project that applies Semantic Web technologies to produce a unified meta-data structure and integrate the training content. As part of this project, a common vocabulary is developed in the form of a *Content Management Ontology*. The ontology's purpose is to provide a knowledge model for describing and linking various multimedia training content (e.g. web pages, documents, presentations, videos) to assist trainers to create new content and reuse existing content, as well as learners to find the relevant content for their training needs. The ontology's scope is restricted to concepts and relationships that describe the main characteristics of training materials developed and used within the organisation, which includes both bespoke training content and content from external sources (e.g. relevant policy documents).

- (a) You have decided to use scenarios as part of the ontology requirements elicitation. Describe appropriate instruments to derive scenarios, considering each of the user groups – tutors and learners.

[4 marks]

- (b) Identify two appropriate knowledge sources which can be consulted when defining the main concepts and relationships of the *Content Management Ontology*. Justify your choice.

[4 marks]

- (c) One way to construct the *Content Management Ontology* is to use tools for learning ontology concepts from a text corpus. Describe one corpus you could use for this purpose and justify your choice [3 marks]. Suggest an approach that could be followed for learning ontology terms and indicate possible challenges [3 marks].

[6 marks]

- (d) Describe an application-based approach to evaluate the domain coverage of the *Content Management Ontology* [3 marks], and compare this approach with a gold standard approach [3 marks].

[6 marks]

[Total Question 1: 20 marks]

Question 2

Assume that you are part of a team which develops a smart career adviser system for university students. It provides an online space to explore different career opportunities, browse example professional profiles, watch video interviews with alumni, and get career development tips. The career adviser uses several ontologies to provide a unified vocabulary defining career-related concepts, such as professions, skills, and qualifications. To provide timely access to career opportunities and professional profiles, your team has decided to use data from the social networking site LinkedIn. The questions below refer to a *Career Profile Ontology* that represents career items included in professional profiles shared on LinkedIn.

- (a) Consider the following core concepts from the *Career Profile Ontology*:

Education
Qualification
Work Experience
Skill

Formulate sentences in the controlled natural language Rabbit to define appropriate OWL axioms to include in the *Career Profile Ontology*. Provide definitions of all classes and two axioms for each of the following: subclasses, disjoint classes, object properties linking concepts, and cardinality restrictions on object properties [2 marks for each]. You may need to define new concepts that are not included in the initial list of core concepts.

[10 marks]

- (b) A colleague of yours asks why you do not use RDF for representing the ontology concepts and relationships. Provide two illustrations, related to Career Profile Ontology, which show natural language statements that can be expressed in OWL but cannot not be expressed in RDF.

[2 marks]

- (c) There are many possible instances of people skills and job roles. Defining each of them manually is highly time consuming. Describe how a crowdsourcing ontology authoring method can be adapted for this task [2 marks]. Identify one challenge of adopting this method for extending the *Career Profile Ontology* [2 marks].

[4 marks]

- (d) Suppose that you have built the *Career Profile Ontology*. You have to evaluate the fitness of the ontology for the purpose of semantically tagging job descriptions. For this, you have conducted a semantic augmentation task over example job descriptions. Describe the main steps you will conduct to validate the output of the semantic augmentation.

[4 marks]

[Total Question 2: 20 marks]

Question 3

This question refers to the hotel ontology developed during the lectures as illustration of the main steps of ontology engineering. The purpose of the ontology was to enable the integration of the datasets between several hotel booking sites. The hotel ontology provides a bootstrapped model of the hotel booking domain including main concepts and relationships related to room booking.

- (a) Draw an RDF graph which presents the meta-data of a hotel reservation. Include the following entities: hotel, customer, reservation duration, payment, and discount. Use appropriate namespaces. Indicate which nodes are classes and which are literals.

[4 marks]

- (b) You need to extend the concepts and relationships in an ontology to include geographic information, e.g. the location of each hotel and the relevant points of interest. For this, you have to find relevant ontologies from the Web of Data. State two criteria that you will consider when selecting data sets and justify why they are important in this case.

[4 marks]

- (c) Your second task is to expose the hotel databases as RDF mapping the fields in the relational databases to ontology concepts. Describe what technologies you will use to conduct this step.

[2 marks]

- (d) You want to extend the instances of your RDF data set to include hotels which are represented in Wikipedia. Hence, you have to conduct SPARQL queries over DBpedia. Consider the following SPARQL query over DBpedia:

```
#prefix declaration
prefix dbp-ont: <http://dbpedia.org/ontology/>

#result clause
SELECT *

#dataset definition
FROM <dbp-ont>

#query pattern
WHERE {
  ?h rdf:type dbp-ont:Hotel.
  ?h dbpprop:location dbpedia:Dubai.
  ?h dbpedia-owl:floorCount ?f.
  ?h rdfs:label ?hotelName.
}
```

Describe the output of this query.

[2 marks]

- (e) The above query is modified by replacing `dbpedia:Dubai` with `dbpedia:United_Arab_Emirates`. The output of the modified query (which refers to United Arab Emirates) has fewer rows than the output of the original query (which refers to Dubai). Describe what can be the reason for this difference.

[3 marks]

- (f) Define a SPARQL query which retrieves from DBpedia a list of all hotels with more than 3 floors. The result of the query should be ordered in an ascending way according to the number of hotel floors and rendered in a three-column table, including the hotel name, hotel location, and the number of hotel floors.

[5 marks]

[Total Question 3: 20 marks]