

This question paper consists
of 6 printed pages each
of which is identified by the Code
Number COMP372501

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

© UNIVERSITY OF LEEDS

School of Computing

January 2017

COMP3725

Knowledge Enriched Information Systems

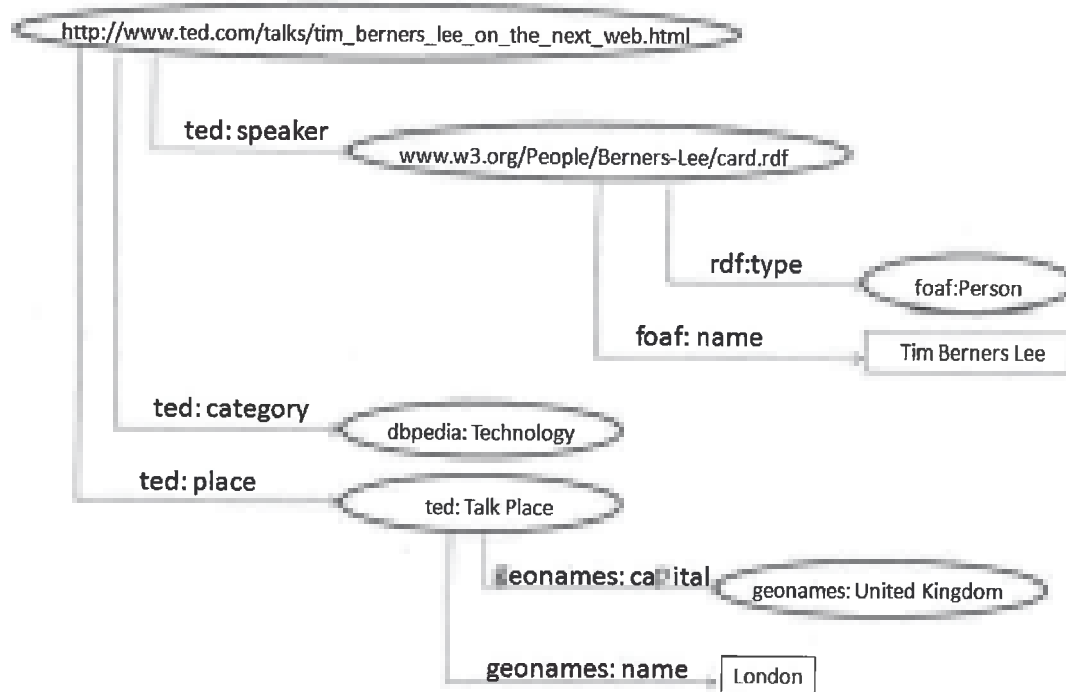
Answer all 3 questions

Time allowed: 2 hours

Question 1

To enable semantic search through the TED (<http://www.ted.com/>) video collection of talks, the metadata about each talk is presented in RDF and linked to knowledge entities, represented in RDF(S) or OWL, describing the speaker and the topic of the talk. The following questions refer to this scenario.

(a) Consider the following RDF graph:



(i) List the namespaces used in the graph, and describe what they represent.

[2 marks]

(ii) Write natural language sentences that state the knowledge shown in the triples of the graph.

[2 marks]

(b) Draw the corresponding triples to add the following facts to the above graph:

- the speaker of this talk knows Jim Hendler (use the property `foaf: knows`);
- United Kingdom is located in Europe (use the property `geonames: part of`).

Note that you should not redraw the graphs but just the triples with appropriate nodes.

[2 marks]

(c) For each of the following propositions, state whether it can be modelled in RDF(S) and justify why:

- (i) Every country is part of a continent.
- (ii) 'part of' is a transitive relationship.
- (iii) Every country has exactly one capital.
- (iv) Paris is the capital of France.

[4 marks]

(d) Consider the following core concepts from an ontology describing videos:

topic
speaker
date
rating
duration

Formulate sentences in the controlled natural language Rabbit to define appropriate axioms for an OWL ontology that includes the above concepts. Provide definitions of all classes and two axioms for each of the following: subclasses (you may need to define new concepts not included in the above list), disjoint classes, object properties linking concepts, and cardinality restrictions on object properties (2 marks for each).

[10 marks]

[Total Question 1: 20 marks]

Question 2

An online publisher of research articles would like to export its meta-data on the Web of Data to enable interoperability and wider deployment. To do so, it has to provide a semantic description of the publications that can be seen in its data set. This will be done using a *Research Articles Ontology* describing the main concepts and relationships of the research articles domain, confined to information that can be seen in the article meta-data (e.g. title, author, publisher, keywords, abstract). The questions below refer to building the *Research Articles Ontology*.

(a) The first step in the *Research Articles Ontology* development is to define appropriate requirements. Define ontology requirements by specifying the intended uses, intended users, main purpose of the ontology, and ontology scope (1 mark for each).

[4 marks]

(b) Before formulating an ontology, it is important to identify appropriate knowledge sources which can be consulted when defining the main concepts and relationships. Identify two knowledge sources for defining the main concepts and relationships in the *Research Articles Ontology* and justify your choice (2 marks for each source).

[4 marks]

(c) A colleague of yours suggests that you use DBpedia – the crowd-sourced community effort allowing extraction of structured information from Wikipedia and making this information available on the Web. Is DBpedia an appropriate knowledge source to use for the *Research Articles Ontology* (2 marks)? Justify your answer (2 marks).

[4 marks]

(d) One way to construct the *Research Articles Ontology* is to use tools for learning ontology concepts from a text corpus. Identify one corpus you could use for this purpose (1 mark) and describe an approach that could be followed for learning ontology terms (2 marks).

[3 marks]

(e) Suppose that you have built the *Research Articles Ontology*. You have to evaluate the fitness of the ontology for the purpose of semantically tagging research articles. For this, you have conducted a semantic augmentation task over the titles and abstracts of a selected set of publications using the Gate system (introduced at lectures). Describe how you will validate the output of the semantic augmentation, including creating a gold standard (3 marks) and evaluation of the semantic annotation (2 marks).

[5 marks]

[Total Question 2: 20 marks]

Question 3

This question refers to DBpedia which extracts RDF triples from the information presented in Wikipedia. Your task is to compose example SPARQL queries to investigate the use of DBpedia as a knowledge source to offer blog writers additional information about things mentioned in their posts. This can stimulate the writer's creativity or extend the information content of their blogs.

(a) Consider the following SPARQL query over the DBpedia ontology:

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

#result clause
SELECT ?eventName ?gp

#dataset definition
FROM <dbp-ont>

#query pattern
WHERE {
  ?e rdf:type dbp-ont:Event.
  ?e rdfs:label ?eventName.
  ?e dbpprop:grandPrix ?gp
}
```

Describe what is the output of this query. Specify how many columns will the resultant table have and what will be the values in each column.

[3 marks]

(b) Modify the query given in (a) to create a new query that lists all events which have status "Active"@en (use the property `dbpprop:status`). The results should be presented in two columns, including the event URI and the event name.

[2 marks]

(c) Define a SPARQL query which returns a list of film festivals from DBpedia. The result should be rendered in a one column table including the event name, ordered in an ascending way according to the event name, and limited to the first 50 entities. Give the whole of the query, including the prefix declaration (1 mark), result clause (1 mark), dataset definition (1 mark), and query pattern (3 marks).

[6 marks]

(d) Describe in natural language the output of the following SPARQL query:

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

#result clause
SELECT *

#dataset definition
FROM <dbp-ont>

#query pattern
WHERE {
  ?s dbpprop:occupation dbpedia:Actor.
  ?s dbp-ont:birthDate ?date.
  ?s dbp-ont:birthPlace dbpedia:Yorkshire.
} ORDER BY (?date)
```

[3 marks]

(e) Follow the example from (d) to define a SPARQL query which returns a list of all composers in DBpedia who were born in England, in a descending order according to their birth date. Give the whole of the query, including the prefix declaration (1 mark), result clause (1 mark), dataset definition (1 mark), and query pattern (3 marks).

[6 marks]

[Total Question 3: 20 marks]