

# CSCI 548 Final Exam (Fall 2014)

---

## Question 1-4: RDF, RDF-Schema, Linked Data and SPARQL (20 points)

Consider the following RDF

```
@prefix : <http://ex.com/soccer> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

:Person a owl:Class .
:Player rdfs:subClassOf :Person .
:Team a owl:Class .

:member a owl:ObjectProperty ;
    rdfs:domain :Team ;
    rdfs:range :Person .

:name a owl:DatatypeProperty .
:description a owl:DataProperty .

:careerPoints a owl:DatatypeProperty ;
    rdfs:domain :Person .

:lakers a :Team ;
    :name "Lakers" ;
    :member [ a :Player ; :name "Bryant" ] ;
    :member :scott .

:scott a :Person ;
    :name "Byron Scott" ;
    :careerPoints 15000 .

[
    :name "LA Clippers" ;
    :member [ :careerPoints 20000 ]
]

[
    :description ""Blake Griffin (born March 16, 1989) is an
American professional basketball player who currently plays for the Los
Angeles Clippers and has scored 7,063 points during his career."" .
]
```

**a)** Draw a graph of the above RDF including a bubble for every class and instance. Use `rdf:type` to indicate the instance/class relationship. Only draw arrows for the statements that are explicitly mentioned in the RDF. (5 points)

**b)** Draw using dotted lines all inferences that can be made using RDF Schema. (2 points)

**c)** The above RDF is not good Linked Data for many reasons. Select two different RDF statements, explain the Linked Data principle it does not follow and why, and rewrite the statement so that it follows the Linked Data principles. The two statements you select should illustrate different problems. (4 points)

**d)** Suppose you load the above RDF in a triple store that implements RDF Schema reasoning. Write a SPARQL query that produces a table with headings as below. The query should return all instances mentioned in the above RDF except for the instance about Blake Griffin". Fill in the cells of the table as appropriate (use as many rows as you need). (5 points)

TeamName	PlayerName	CareerPoints

Consider the following query

```
select ?x {
    ?x a :Person .
}
```

**e)** If the triple store supports RDF Schema reasoning, what will this query return? (2 points)

**f)** If the triple store supports NO RDF Schema reasoning, what will this query return? (2 points)

### Question 5: Ontology-based Rewriting (5pts)

Produce the perfect rewriting of query:  $q(x) \leftarrow \text{Student}(x) \wedge \text{enrolled}(x,y)$   
under the following ontology:

o1:  $\text{GraduateStudent} \subseteq \text{Student}$

o2:  $\exists \text{enrolled}^- \subseteq \text{Course}$

Please show your derivation steps.

### Question 6: Big Data (5pts)

Not applicable

### Question 7: Data Cleaning (5pts)

Please describe the steps used to transform the original table to the target table.

Original Table:

Name	Major	Score
Ann	Math	43%
Ann	Bio	78%
Bob	Math	96%
Bob	Bio	54%

Target Table:

Name	Math	Bio
Ann	43	78
Bob	96	54

### Question 8: Query Containment (5pts)

Given Q1 & Q2:

Q1:  $p(X,Y) :- r(X,Y) \ \& \ X=Y$

Q2:  $p(A,B) :- r(A,B) \ \& \ r(A,C)$

Is Q1 contained in Q2? If yes, give the containment mapping.

Is Q2 contained in Q1? If yes, give the containment mapping.

### Question 9: Answering Queries Using Views (5pts)

Name and describe one algorithm for doing query reformulation using LAV rules.

### Question 10: Schema Mapping (5pts)

True or false:

- 1) A schema matcher can only discover 1-1 matches
- 2) Schema mapping is the first step in the larger schema matching problem

### Question 11: Automatic Source Modeling (5pts)

In the work on Learning Semantic Descriptions of Web Information Sources (by Carman), how does the system determine whether it has correctly learned a description of a data source?

### Question 12: Karma (5pts)

Karma is to \_\_\_\_\_ like Marmite is to HTML

Karma is to \_\_\_\_\_ like R2RML is to RDF

Give an example where it is important to represent relationships among semantic types. Explain why ( $\leq 2$  sentences)

### Question 13: OWL2 (5pts)

A HappyMother is a female human who has a son

A HappyChild is a young human whose parents are alive and happy

Use OWL2 axioms to precisely define a HappyMother. You can use the classes Human, Male, Female, Happy, Alive, Young and the properties hasChild, hasParent

Use OWL2 axioms to precisely define a HappyChild. You can use the classes Human, Male, Female, Happy, Alive, Young and the properties hasChild, hasParent

### Question 14: String Matching (5pts)

Dataset 1	Dataset 2
Brad Pitt	Pitt, Brad
Angelina Jolie	Jolie, Angelina

Which metric would you use to compute similarity between dataset 1 and dataset 2? Explain.

Dataset 1	Dataset 2
Brad Pitt	brad pit
Angelina Jolie	angelin joule

Which metric would you use to compute similarity between dataset 1 and dataset 2? Explain.



**(1) Question 15: Record Matching (5pts)**

Briefly describe one advantage of learning-based approaches to record linkage over rule-based approaches.

**Question 16: Mashups (5pts)**

(1) Explain Intel Mashmaker's "Pay as you go approach" in regards to basic and expert users

(2) Explain one of the biggest problems users experienced with Marmite.

### Question 17: Information Extraction (5pts)

Given text such as “\$25 for **The Hobbit: The Battle of the Five Armies** at **AMC, Santa Monica** at **9pm on Dec. 17**” on Web sites, and suppose you are interested in extracting information of ticket prices, movie names, theaters, location and show times” from the text. However, there is only one available reference set containing only information of movie names and theater names such as in the following table. What’s your solution to extract information of prices, locations and show times?

Movie	Theater
The Hobbit: The Battle of the Five Armies	AMC
Hunger Game: Mockingjay Part I	Regal

### Question 18: Wrapper Learning (5pts)

Please briefly describe the following types of wrappers

- (1) LR
- (2) HLRT
- (3) BELR

### Question 19: Wrapper Generation (5pts)

What are the three steps used by AutoFeed to generate relation tables? Please briefly describe these three steps.

### Question 20: Intellectual Property (5pts)

You invent a new web-based user interface for a search engine that provides a much improved user experience and will transform the way users interact with a search engine. What method should you use to protect this user interface? Why?