M2R Exam – Semantic web: from XML to OWL
Social and semantic web part

Duration : 2h00
Any document allowed – no communication device allowed

January 2014

Note: Please, carefully read all the questions before answering.

Recommendations

Consider the following expression:

$$predictedRating(u,i) = \Sigma_{i' \in I} ItemSim(i,i') \times rating(u,i')$$

1. Describe an example $ItemSim(i,i')$ function discussed in class.
2. Provide a general formula based on user similarities for predicting item ratings (collaborative filtering)
3. Provide or describe in text two examples of the formula in 2. on real datasets.

Social Top-K Processing

Top-K processing algorithms rely on a pruning condition also called threshold condition. Consider a query $Q = t_1\ t_2$ and the following two scoring functions on Delicious datasets:

$$score1(u,i,t_i) = |Network(u) \cap taggers(i,t_i)|$$
$$score2(u,i,Q) = score(u,i,t_1) - score(u,i,t_2)$$
$$where\ \ score(u,i,Q) = \Sigma_{t_i \in Q} score(u,i,t_i)$$

1. Explain *score1()* and give a formulation and an explanation of $Network(u)$ in Delicious.
2. Which one of the two scoring functions would you use for efficient processing and why?
3. How is efficiency measured for Fagin-style algorithms for top-k processing?
4. Describe in detail an algorithm of your choice No-Random-Access (NRA) or Threshold Algorithm (TA) to find the top-1 answer to Q, for 2 users and 3 items.
5. What is the difference between modeling single-user recommendation and group-recommendation?

User Studies and Relevance Assessment

1. What is the benefit of using Amazon Mechanical Turk instead of traditional user studies?
2. How do we avoid hiring the wrong workers in Amazon Mechanical Turk? Give an example.
3. Define CG, DCG and nDCG in your own terms and develop an example with 3 documents.

RDF Manipulation

Consider the following statements composing the RDF graph G:

```
ex:book1 rdf:type mr:Book .
ex:book1 dc:title "For whom the bell tolls" .
ex:book1 dc:date 1940 .
ex:book1 dc:creator ex:eh .
ex:book1 mr:storedIn "Living room" .

ex:book2 rdf:type mr:Book .
ex:book2 dc:title "Pour qui sonne le glas" .
ex:book2 mr:translationOf ex:book1 .
ex:book2 dc:creator ex:eh .
ex:book2 dc:date 1948 .
ex:book2 dc:publisher ex:gal .

ex:movie1 rdf:type mr:Movie .
ex:movie1 dc:title "For whom the bell tolls" .
ex:movie1 mr:adaptedFrom ex:book1 .
ex:movie1 mr:director ex:sw .
ex:movie1 mr:cast ex:ib .
ex:movie1 mr:cast ex:gc .
ex:movie1 mr:storedIn "Computer drive" .
ex:movie1 dc:date 1943 .

ex:eh rdf:type foaf:Person .
ex:eh foaf:name "Ernest Hemingway" .

ex:sw rdf:type foaf:Person .
ex:sw foaf:name "Sam Wood .

ex:ib rdf:type foaf:Person .
ex:ib foaf:name "Ingrid Bergman" .

ex:gc rdf:type foaf:Person .
ex:gc foaf:name "Gary Cooper" .

ex:gal rdf:type foaf:Organization .
ex:gal foaf:name "Gallimard" .
```

1. Which different classes are there in graph G? Why are they classes?
2. Express G as a graph in graphical form.

Consider the following as the RDF graph G' (`_:` are blank identifiers):

```
_:x rdf:type mr:Movie .
_:x mr:adaptedFrom _:y .
_:y dc:creator _:p .
_:p foaf:name "Ernest Hemingway" .
```

3. Paraphrase the graph G'.
4. Is G' entailed by G? Detail why?

FRBR in RDFS

FRBR is now a well established vocabulary in libraries. FRBR distinguishes between:

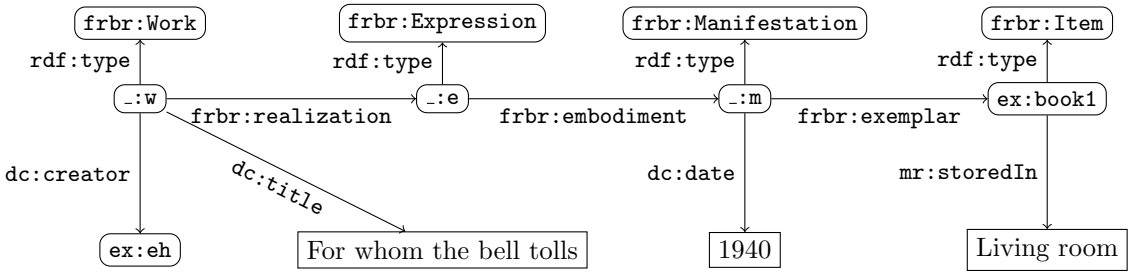
- a *work* which is an abstract idea of what a work is;
- an *expression* which is a realization of this work in a particular form (poetry, music, painting);
- a *manifestation* which is a distinct embodiment of the expression (an edition of a text, a release of a movie; the reproduction of a painting);
- an *item* which is an often physical exemplar of a manifestation (an exemplar of a book; a copy of an MP3 file).

Consider a fragment S of its RDF Schema manifestation:

```
frbr:Work rdf:type rdfs:Class .
frbr:Expression rdf:type rdfs:Class .
frbr:Manifestation rdf:type rdfs:Class .
frbr:Item rdf:type rdfs:Class .
frbr:Performance rdfs:subClassOf frbr:Expression .

frbr:adaptation rdfs:domain frbr:Work .
frbr:adaptation rdfs:range frbr:Work .
frbr:realization rdfs:domain frbr:Work .
frbr:realization rdfs:range frbr:Expression .
frbr:translation rdfs:domain frbr:Expression .
frbr:translation rdfs:range frbr:Expression .
frbr:embodiment rdfs:domain frbr:Expression .
frbr:embodiment rdfs:range frbr:Manifestation .
frbr:exemplar rdfs:domain frbr:Manifestation .
frbr:exemplar rdfs:range frbr:Item .
```

For instance, the first 5 statements of G in FRBR would correspond to the following graph:



5. How would you extend the above S with the vocabulary identified by the prefix `mr` in graph G? I.e., give triples using the RDFS vocabulary (`rdfs:subClassOf`, `rdfs:subPropertyOf`, `rdfs:domain`, `rdfs:range`) to extend FRBR with the corresponding entities.

Refactoring graphs with SPARQL CONSTRUCT

6. Both movies and books may be considered as work expressions. Write a SPARQL query able to return all such items in graph G with their title and, if possible, the place they are stored in.

7. Consider that, instead of extending the schema, one would prefer to refactor the graph G so that it corresponds to the schema S. Create a SPARQL CONSTRUCT query able to extract the data from G and generate a graph complying to S (check on the example above).

8. Considering that you have a SPARQL engine able to answer queries taking into account RDFS semantics, write the same query as in Question 6 using the schema S and the answer to Question 5.