M2R Exam – Semantic web: from XML to OWL
Semantic web part

Duration : 1h30
Any document allowed – no communication device allowed

January 2011

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

RDF and ontologies

Here are the 8 triples of an RDF graph G about writers and their works: (all identifiers correspond in fact to URIs, _:b is a blank node):

$\langle$`d:Poe`, `o:wrote`, `d:TheGoldBug`$\rangle$ $\langle$`d:Baudelaire`, `o:translated`, `d:TheGoldBug`$\rangle$

$\langle$`d:Poe`, `o:wrote`, `d:TheRaven`$\rangle$ $\langle$`d:Mallarmé`, `o:translated`, `d:TheRaven`$\rangle$

$\langle$`d:TheRaven`, `rdf:type`, `o:Poem`$\rangle$ $\langle$`d:Mallarmé`, `o:wrote`, `_:b`$\rangle$

$\langle$`_:b`, `rdf:type`, `o:Poem`$\rangle$ $\langle$`d:TheGoldBug`, `rdf:type`, `o:Novel`$\rangle$

1. Draw an RDF graph corresponding to these statements

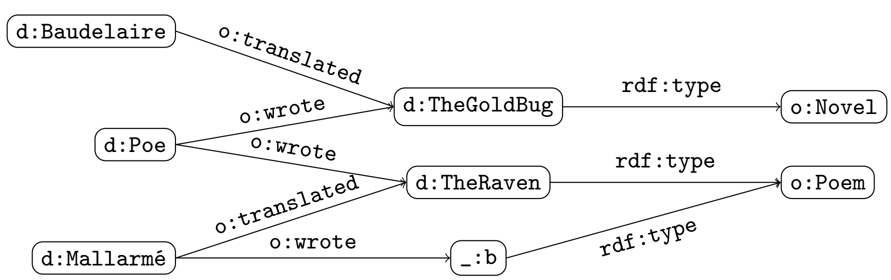

Figure 1: The RDF graph G.

2. Express in English the meaning of these statements.

   Poe wrote the Poem "The Raven" translated by Mallarmé and the novel "The Gold Bug" translated by Baudelaire. Mallarmé wrote a poem.

Consider the RDFS ontology o containing, in addition to those of G, the following statements:

$\langle$`o:Novel`, `rdfs:subClassOf`, `o:Literature`$\rangle$

$\langle$`o:Poem`, `rdfs:subClassOf`, `o:Literature`$\rangle$

$\langle$`o:translated`, `rdfs:range`, `o:Literature`$\rangle$

$\langle$`o:wrote`, `rdfs:domain`, `o:Writer`$\rangle$

3. Does this allow to conclude that `d:Poe`, `d:Baudelaire` or `d:Mallarmé` is a `o:Writer`? Explain why.

   The only assertion that would allow to conclude that someone is a `o:Writer` is the last one related to the domain of the `o:wrote` predicate. Nothing allows for inferring triples with the `o:wrote` predicate, so the only assertions with it are those asserted. Hence, the only writers are `d:Poe` and `d:Mallarmé`.

4. Can you express in OWL the statement that "anyone who write Literature is a Writer"?

   The sentence expresses that those who write Literature are Writers, hence Writer is a superclass of the restriction. This can be expressed by creating a class equivalent to the restriction, and subclass of Writer:

```
<owl:Class>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource="o:wrote"/>
      <owl:someValueFrom rdf:resource="o:Literature"/>
    </owl:Restriction>
  </owl:equivalentClass>
  <rdfs:subClassOf rdf:resource="o:Writer"/>
</owl:Class>
```

   or more briefly:

```
<owl:Restriction>
  <owl:onProperty rdf:resource="o:wrote"/>
  <owl:someValueFrom rdf:resource="o:Literature"/>
  <rdfs:subClassOf rdf:resource="o:Writer"/>
</owl:Restriction>
```

   It would have been possible to add a range constraint on the `o:wrote` predicate so that whatever is written is `o:Literature` ($\langle$`o:wrote`, `rdfs:range`, `o:Literature`$\rangle$). However, this is stronger than what was asked since it would have restricted a particular author to write something else than literature.

SPARQL query containment

Consider the following queries q_1 and q_2 on the RDF graph of the previous exercise:

- $q_1 =$ `SELECT` $?w$ `FROM` G `WHERE` $(\langle ?w\ \texttt{o:wrote}\ ?x\rangle$ `AND` $\langle ?x\ \texttt{rdf:type}\ \texttt{o:Poem}\rangle)$ `UNION` $\langle ?w\ \texttt{o:translated}\ ?x\rangle$;
- $q_2 =$ `SELECT` $?w$ `FROM` G `WHERE` $(\langle ?w\ \texttt{o:wrote}\ ?l\rangle$ `UNION` $\langle ?w\ \texttt{o:translated}\ ?l\rangle)$ `AND` $\langle ?l\ \texttt{rdf:type}\ \texttt{o:Poem}\rangle$.

5. In the course, we defined the distinguished variables $\vec{B}$, the queried graph G and the query pattern P. Identify them in q_1 and q_2.

   In both cases, $\vec{B}$ is $\langle ?w\rangle$ and G is G. The patterns of q_1 and q_2 are respectively :

$$P_1 = (\langle ?w\ \texttt{o:wrote}\ ?x\rangle\ \texttt{AND}\ \langle ?x\ \texttt{rdf:type}\ \texttt{o:Poem}\rangle)\ \texttt{UNION}\ \langle ?w\ \texttt{o:translated}\ ?x\rangle$$
$$P_2 = (\langle ?w\ \texttt{o:wrote}\ ?l\rangle\ \texttt{UNION}\ \langle ?w\ \texttt{o:translated}\ ?l\rangle)\ \texttt{AND}\ \langle ?l\ \texttt{rdf:type}\ \texttt{o:Poem}\rangle$$

6. Provide the answers of q_1 and q_2 with respect to the graph G.

   The answers to query q_1 and q_2 on the graph G are respectively $\{\langle$`d:Poe`$\rangle, \langle$`d:Mallarmé`$\rangle, \langle$`d:Baudelaire`$\rangle\}$ and $\{\langle$`d:Poe`$\rangle, \langle$`d:Mallarmé`$\rangle\}$.

Query containment $q \sqsubseteq q'$ between two queries $q =$ `SELECT` $\vec{B}$ `FROM` G `WHERE` P and $q' =$ `SELECT` $\vec{B}$ `FROM` G `WHERE` P' is defined by the fact that for any RDF graph, the answers to q are included in those to q' ($\forall G, \mathcal{A}(\vec{B}, G, P) \subseteq \mathcal{A}(\vec{B}, G, P')$).

7. What does the answer to the previous questions tell you about query containment between q_1 and q_2?

   $q_1 \not\sqsubseteq q_2$ because G is a counter example: $\{\langle \texttt{d:Poe}\rangle, \langle \texttt{d:Mallarmé}\rangle, \langle \texttt{d:Baudelaire}\rangle\} = \mathcal{A}(\langle ?w\rangle, G, P_1) \not\subseteq \mathcal{A}(\langle ?w\rangle, G, P_2) = \{\langle \texttt{d:Poe}\rangle, \langle \texttt{d:Mallarmé}\rangle\}$.

8. Do you think that query containment holds in some direction between q_1 and q_2 (either $q_1 \sqsubseteq q_2$ or $q_2 \sqsubseteq q_1$)?

   $q_2 \sqsubseteq q_1$.

9. Provide a proof for this. This may be done semantically by using the interpretation of query patterns or syntactically by translating queries into logic and showing that the query containment statement is a theorem.

   The argument for the proof is that $P_1 = (A \wedge B) \vee C$ and $P_2 = (A \vee C) \wedge B$, but $(A \vee C) \wedge B = (A \wedge B) \vee (C \wedge B)$. Hence, P_2 is more specific than P_1. More formally:

$$
\begin{aligned}
& \sigma \in A(\langle ?w\rangle, G, P_2) \\
\Leftrightarrow\ & G \models \sigma((\langle ?w \texttt{ o:wrote } ?l\rangle \texttt{ UNION } \langle ?w \texttt{ o:translated } ?l\rangle) \texttt{ AND } \langle ?l \texttt{ rdf:type o:Poem}\rangle) \\
\Leftrightarrow\ & G \models \sigma(\langle ?w \texttt{ o:wrote } ?l\rangle \texttt{ UNION } \langle ?w \texttt{ o:translated } ?l\rangle) \text{ and } G \models \sigma(\langle ?l \texttt{ rdf:type o:Poem}\rangle) \\
\Leftrightarrow\ & (G \models \sigma(\langle ?w \texttt{ o:wrote } ?l\rangle) \text{ or } G \models \sigma(\langle ?w \texttt{ o:translated } ?l\rangle)) \text{ and } G \models \sigma(\langle ?l \texttt{ rdf:type o:Poem}\rangle) \\
\Leftrightarrow\ & (G \models \sigma(\langle ?w \texttt{ o:wrote } ?l\rangle) \text{ and } G \models \sigma(\langle ?l \texttt{ rdf:type o:Poem}\rangle)) \\
& \text{ or } (G \models \sigma(\langle ?w \texttt{ o:translated } ?l\rangle) \text{ and } G \models \sigma(\langle ?l \texttt{ rdf:type o:Poem}\rangle)) \\
\Rightarrow\ & (G \models \sigma(\langle ?w \texttt{ o:wrote } ?l\rangle) \text{ and } G \models \sigma(\langle ?l \texttt{ rdf:type o:Poem}\rangle)) \text{ or } G \models \sigma(\langle ?w \texttt{ o:translated } ?l\rangle) \\
\Leftrightarrow\ & G \models \sigma(\langle ?w \texttt{ o:wrote } ?x\rangle \texttt{ AND } \langle ?x \texttt{ rdf:type o:Poem}\rangle) \text{ or } G \models \sigma(\langle ?w \texttt{ o:translated } ?x\rangle) \\
\Leftrightarrow\ & G \models \sigma((\langle ?w \texttt{ o:wrote } ?x\rangle \texttt{ AND } \langle ?x \texttt{ rdf:type o:Poem}\rangle) \texttt{ UNION } \langle ?w \texttt{ o:translated } ?x\rangle) \\
\Leftrightarrow\ & \sigma \in A(\langle ?w\rangle, G, P_1)
\end{aligned}
$$

   Hence, $q_2 \sqsubseteq q_1$.

Query modulo ontology

We now consider the ontology o and the following queries:

- $q_3 = \texttt{SELECT } ?y \texttt{ FROM } o \texttt{ WHERE } \langle ?x, \texttt{o:translated}, ?y\rangle$;
- $q_4 = \texttt{SELECT } ?y \texttt{ FROM } o \texttt{ WHERE } \langle ?y, \texttt{rdf:type}, \texttt{o:Literature}\rangle$.

10. Do you think that query containment holds in some direction between q_3 and q_4 (either $q_3 \sqsubseteq q_4$ or $q_4 \sqsubseteq q_3$)? Tell why.

    None of these because SPARQL evaluates queries by finding triples in the graph and the triples are not comparable. More formally, assume $G_1 = \{\langle a, \texttt{o:translated}, b\rangle\}$ and $G_2 = \{\langle c, \texttt{rdf:type}, \texttt{o:Literature}\rangle\}$, it is clear that $\mathcal{A}(q_3, G_1) \not\subseteq \mathcal{A}(q_4, G_1)$ and $\mathcal{A}(q_4, G_2) \not\subseteq \mathcal{A}(q_3, G_2)$. Hence, there cannot be any containment between these queries.

11. Can you provide a definition for query containment modulo an ontology o ($q \sqsubseteq_o q'$)?

    There is no reason to change the structure of the definition: Query containment $q \sqsubseteq q'$ between two queries $q = \texttt{SELECT } \vec{B} \texttt{ FROM } o \texttt{ WHERE } P$ and $q' = \texttt{SELECT } \vec{B} \texttt{ FROM } o \texttt{ WHERE } P'$ is defined by the fact that for any RDFS ontologies, the answers to q are included in those to q' ($\forall o, \mathcal{A}^+(\vec{B}, o, P) \subseteq \mathcal{A}^+(\vec{B}, o, P')$).

    Everything is in the definition of $\mathcal{A}^+$. A natural semantic definition would be that:

$$\mathcal{A}^+(\vec{B}, o, P) = \{\sigma|_{\vec{B}}^{\vec{B}} | o \models_{RDFS} \sigma(P)\}$$

    or a more pragmatic approach would be to define it with the closure:

$$\mathcal{A}^+(\vec{B}, o, P) = \mathcal{A}(\vec{B}, \hat{o}\backslash\backslash P, P)$$

12. Does it return different answers for q_3 and q_4?(either $q_3 \sqsubseteq_o q_4$ or $q_4 \sqsubseteq_o q_3$)? Tell why.

    From the definition of o, it is clear that whenever $o \models_{RDFS} \langle a, \texttt{o:translated}, b\rangle$, then $o \models_{RDFS} \langle b, \texttt{rdf:type}, \texttt{o:Literature}\rangle$. The converse is not true (there exists models satisfying $\langle b, \texttt{rdf:type}, \texttt{o:Literature}\rangle$ but no a such that $\langle a, \texttt{o:translated}, b\rangle$, i.e., there may be non translated books). Then for any $\langle b\rangle \in \mathcal{A}^+(q_3)$, $\langle b\rangle \in \mathcal{A}^+(q_4)$, so $q_3 \sqsubseteq_o q_4$. But not the other way around.

Network of ontologies

We now consider an ontology o' which defines the class `op:Buch` and contains the following statements:

$$\langle \texttt{d:Baudelaire}, \texttt{o:translated}, \texttt{d:Confessions}\rangle \ \langle \texttt{d:DeQuincey}, \texttt{o:wrote}, \texttt{d:Confessions}\rangle$$

and o'' which defines the class `opp:Roman` and contain the following statements:

$$\langle \texttt{d:Confessions}, \texttt{rdf:type}, \texttt{opp:Roman}\rangle \ \langle \texttt{d:Musset}, \texttt{o:translated}, \texttt{d:Confessions}\rangle$$

They are related together by the following three alignments:

- $A_{o,o'} = \{\langle \texttt{o:Literature}, \equiv, \texttt{op:Buch}\rangle\}$
- $A_{o',o''} = \{\langle \texttt{op:Buch}, \sqsubseteq, \texttt{opp:Roman}\rangle\}$
- $A_{o'',o} = \{\langle \texttt{opp:Roman}, \equiv, \texttt{o:Novel}\rangle\}$

So that we have a network of ontology $\langle\{o, o', o''\}, \{A_{o,o'}, A_{o',o''}, A_{o'',o}\}\rangle$.

13. Do you think that this network of ontologies is well designed? Why?

    It is correctly defined because it is made a set of ontologies and a set of alignments between these ontologies. However, the statement $\texttt{op:Buch} \sqsubseteq \texttt{opp:Roman}$ seems strange and maybe exactly the opposite.

14. Is this network consistent? Provide a model for this network of ontologies.

    The network is indeed consistent. As a model it is possible to create a model isomorphic to the ontologies (with, in each of the ontologies, the same URI interpreted in the same way and equivalent classes having the same interpretation). A model of the network may have been a triple $\langle m, m', m''\rangle$ such that:

$$\begin{aligned}
m(\texttt{o:Literature}) = m'(\texttt{op:Buch}) = m''(\texttt{opp:Roman}) &= m(\texttt{o:Novel})\\
m(\texttt{op:Poem}) &\subseteq m(\texttt{o:Literature})\\
\{m(\texttt{d:Poe}), m(\texttt{d:Mallarmé}), m'(\texttt{d:DeQuincey})\} &\subseteq m(\texttt{Writer})\\
\{m''(\texttt{d:Confessions}), m(\texttt{d:TheGoldBug})\} &\subseteq m(\texttt{o:Literature})\\
\{m(\texttt{d:TheRaven}), m(\texttt{d:Brise marine})\} &\subseteq m(\texttt{o:Poem})\\
m(\texttt{_:b}) &= m(\texttt{d:Brise marine})\\
\langle m(\texttt{d:Poe}), m(\texttt{d:TheGoldBug})\rangle &\in m(\texttt{o:wrote})\\
\langle m(\texttt{d:Poe}), m(\texttt{d:TheRaven})\rangle &\in m(\texttt{o:wrote})\\
\langle m(\texttt{d:Mallarmé}), m(\texttt{d:Brise Marine})\rangle &\in m(\texttt{o:wrote})\\
\langle m'(\texttt{d:DeQuincey}), m'(\texttt{d:Confessions})\rangle &\in m'(\texttt{o:wrote})\\
\langle m(\texttt{d:Mallarmé}), m(\texttt{d:TheRaven})\rangle &\in m(\texttt{o:translated})\\
\langle m(\texttt{d:Baudelaire}), m(\texttt{TheGoldBug})\rangle &\in m(\texttt{o:translated})\\
\langle m''(\texttt{d:Musset}), m''(\texttt{d:Confessions})\rangle &\in m''(\texttt{o:translated})\\
\langle m'(\texttt{d:Baudelaire}), m'(\texttt{d:Confessions})\rangle &\in m'(\texttt{o:translated})
\end{aligned}$$

15. Provide the constraints that the alignments impose on models.

    The constraints are that:

$$\begin{aligned} m(\texttt{o:Literature}) &= m'(\texttt{op:Buch}) \\ m'(\texttt{op:Buch}) &\subseteq m''(\texttt{opp:Roman}) \\ m''(\texttt{opp:Roman}) &= m(\texttt{o:Novel}) \end{aligned}$$

    but since $m(\texttt{o:Novel}) \subseteq m(\texttt{o:Literature})$, we have $m(\texttt{o:Literature}) = m'(\texttt{op:Buch}) = m''(\texttt{opp:Roman}) = m(\texttt{o:Novel})$.

16. What does this entail for the class (`rdf:type`) of `d:Confessions` and `d:TheRaven` at o in this network?

    This entails that both works have all these four classes as `rdf:type`. In particular, $\langle$`d:TheRaven`, `rdf:type`, `o:Poem`$\rangle$ and $\langle$`d:TheRaven`, `rdf:type`, `o:Novel`$\rangle$.