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):

$$
\begin{align*}
(d:\text{Poe}, o:\text{wrote}, d:\text{TheGoldBug}) & (d:\text{Baudelaire}, o:\text{translated}, d:\text{TheGoldBug}) \\
(d:\text{Poe}, o:\text{wrote}, d:\text{TheRaven}) & (d:\text{Mallarmé}, o:\text{translated}, d:\text{TheRaven}) \\
(d:\text{TheRaven}, \text{rdf:type}, o:\text{Poem}) & (d:\text{Mallarmé}, o:\text{wrote}, _:_b) \\
(_:_b, \text{rdf:type}, o:\text{Poem}) & (d:\text{TheGoldBug}, \text{rdf:type}, o:\text{Novel})
\end{align*}
$$

1. Draw an RDF graph corresponding to these statements

![RDF graph](image)

Figure 1: The RDF graph $G$.

2. Express in English the meaning of these statements.


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

$$
\begin{align*}
(o:\text{Novel}, \text{rdfs:subClassOf}, o:\text{Literature}) \\
(o:\text{Poem}, \text{rdfs:subClassOf}, o:\text{Literature}) \\
(o:\text{translated}, \text{rdfs:range}, o:\text{Literature}) \\
(o:\text{wrote}, \text{rdfs:domain}, o:\text{Writer})
\end{align*}
$$
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 (⟨o:wrote, rdfs:range, o:Literature⟩). 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 ((?w o:wrote ?x) AND (?x rdf:type o:Poem)) UNION (?w o:translated ?x);

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 (?w) and $G$ is G. The patterns of $q_1$ and $q_2$ are respectively:

- $P_1 =$ (?w o:wrote ?x) AND (?x rdf:type o:Poem)) UNION (?w o:translated ?x)
- $P_2 =$ (?w o:wrote ?l) UNION (?w o:translated ?l)) AND (?l rdf:type o:Poem)

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 {⟨d:Poe⟩, ⟨d:Mallarmé⟩, ⟨d:Baudelaire⟩} and {⟨d:Poe⟩, ⟨d:Mallarmé⟩}.

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 $q'$ ($\forall G, A(\vec{B}, G, P) \subseteq A(\vec{B}, G, P')$).
7. What does the answer to the previous questions tell you about query containment between \( q_1 \) and \( q_2 \)?

\( q_1 \nsubseteq q_2 \) because \( G \) is a counter example: \( \{\langle \text{d:Poe}, \text{d:Mallarmé}, \text{d:Baudelaire} \rangle \} = \mathcal{A}(?w), G, P_1 \nsubseteq \mathcal{A}(?w), G, P_2 = \{\langle \text{d:Poe}, \text{d:Mallarmé} \rangle \} \).

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

\( q_2 \subseteq q_1 \).

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 \( P_1 = (A \land B) \lor C \) and \( P_2 = (A \lor C) \land B \), but \( (A \lor C) \land B = (A \land B) \lor (C \land B) \). Hence, \( P_2 \) is more specific than \( P_1 \). More formally:

\[
\sigma \in \mathcal{A}(?w), G, P_2 \iff G \models \sigma(\langle ?w \text{ o:translated } ?l \rangle \text{ UNION } \langle ?l \text{ rdf:type o:Poem} \rangle) \\
\quad \iff G \models \sigma(\langle ?w \text{ o:translated } ?l \rangle \text{ UNION } \langle ?l \text{ rdf:type o:Poem} \rangle) \text{ and } G \models \sigma(\langle ?l \text{ rdf:type o:Poem} \rangle)
\]

Hence, \( q_2 \subseteq q_1 \).

**Query modulo ontology**

We now consider the ontology \( o \) and the following queries:

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

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

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

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

There is no reason to change the structure of the definition: Query containment \( q \subseteq q' \) between two queries \( q = \text{SELECT } \vec{B} \text{ FROM } o \text{ WHERE } P \) and \( q' = \text{SELECT } \vec{B} \text{ FROM } o \text{ WHERE } P' \) 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:

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

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

\[
\mathcal{A}^+(\vec{B}, o, P) = \mathcal{A}(\vec{B}, o) \setminus P, P)
\]
12. Does it return different answers for $q_3$ and $q_4$? (either $q_3 \sqsubseteq o \sqsubseteq q_4$ or $q_4 \sqsubseteq o \sqsubseteq q_3$)? Tell why.

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

Network of ontologies

We now consider an ontology $o'$ which defines the class $\text{op:Buch}$ and contains the following statements:

$$\langle d; \text{Baudelaire}, o; \text{translated}, d; \text{Confessions} \rangle, \langle d; \text{DeQuincey}, o; \text{wrote}, d; \text{Confessions} \rangle$$

and $o''$ which defines the class $\text{opp:Roman}$ and contain the following statements:

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

They are related together by the following three alignments:

$$A_{o,o'} = \{\langle o; \text{Literature}, \sqsubseteq, \text{op:Buch} \rangle\}$$

$$A_{o',o''} = \{\langle \text{op:Buch}, \sqsubseteq, \text{opp:Roman} \rangle\}$$

$$A_{o'',o} = \{\langle \text{opp:Roman}, \sqsubseteq, o; \text{Novel} \rangle\}$$

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

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 $\text{op:Buch} \sqsubseteq \text{opp:Roman}$ seems strange and maybe exactly the opposite.


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:

$$m(o; \text{Literature}) = m'(\text{op:Buch}) = m''(\text{opp:Roman}) = m(o; \text{Novel})$$

$$m(\text{op:Poem}) \sqsubseteq m(o; \text{Literature})$$

$$\{m(d; \text{Poe}), m(d; \text{Mallarmé}), m'(d; \text{DeQuincey})\} \subseteq m(\text{Writer})$$

$$\{m''(d; \text{Confessions}), m(d; \text{TheGoldBug})\} \subseteq m(o; \text{Literature})$$

$$\{m(d; \text{TheRaven}), m(d; \text{Brise marine})\} \subseteq m(o; \text{Poem})$$

$$m(\_; b) = m(d; \text{Brise marine})$$

$$\langle m(d; \text{Poe}), m(d; \text{TheGoldBug})\rangle \in m(o; \text{wrote})$$

$$\langle m(d; \text{Poe}), m(d; \text{TheRaven})\rangle \in m(o; \text{wrote})$$

$$\langle m(d; \text{Mallarmé}), m(d; \text{Brise Marine})\rangle \in m(o; \text{wrote})$$

$$\langle m'(d; \text{DeQuincey}), m'(d; \text{Confessions})\rangle \in m'(o; \text{wrote})$$

$$\langle m(d; \text{Mallarmé}), m(d; \text{TheRaven})\rangle \in m(o; \text{translated})$$

$$\langle m(d; \text{Baudelaire}), m(d; \text{TheGoldBug})\rangle \in m(o; \text{translated})$$

$$\langle m''(d; \text{Musset}), m''(d; \text{Confessions})\rangle \in m''(o; \text{translated})$$

$$\langle m'(d; \text{Baudelaire}), m'(d; \text{Confessions})\rangle \in m'(o; \text{translated})$$
15. Provide the constraints that the alignments impose on models.

The constraints are that:

\[
\begin{align*}
m(\text{o:Literature}) &= m'(\text{op:Buch}) \\
m'(\text{op:Buch}) &\subseteq m''(\text{opp:Roman}) \\
m''(\text{opp:Roman}) &= m(\text{o:Novel})
\end{align*}
\]

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

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

This entails that both works have all these four classes as \text{rdf:type}. In particular, \(\langle \text{d:TheRaven}, \text{rdf:type}, \text{o:Poem} \rangle\) and \(\langle \text{d:TheRaven}, \text{rdf:type}, \text{o:Novel} \rangle\).