Web of Data - exam*

Documents allowed – no communication device allowed – duration: 2:30

Computer Science Master 2 – Data Science – Paris Saclay University

15^{th} of January, 2024

1 Part 1: Data Linking (9pts)

- *Question 1 (2 pts).* Give three main families of data linking approaches and, for each, give its main characteristics.
- *Question 2. (2 pts)* What are the main aspects that may be considered for evaluating data linking approaches.
- Let us consider two datasets D_1 and D_2 shown in table 1 and 2 which give an extract of some film descriptions. These films are described by five properties {title, hasActor, rDate, director, lang}.
  We note that the properties **hasActor*** and **director*** are multi-valued and we consider that for each pair of equal values we have a starting $synVals$:

```
synVals("Ocean's 11", "Ocean's 11"), synVals("2004", "2004"),
synVals("P. Greengrass", "P. Greengrass"), synVals("J. Roberts", "J. Roberts"),
...
```

	title	**hasActor***	**rDate**	**director***	**lang**
i_1	Ocean's 11	J. Roberts; B. Pitt;	2001	S. Soderbergh	
i_2	Ocean's 12	J. Roberts; B. Pitt; G. Clooney	2004	S. Soderbergh; P.Greengrass	
i_3	Ocean's 13	B. Pitt; G. Clooney	2007	S. Soderbergh	
i_4	The descendants	N. Krause; G. Clooney	2011	A. Payne	en
i_5	Bourne Identity		2002	P. Greengrass	en
i_6	Ocean's twelve	J. Roberts; B. Pitt; G. Clooney	2004		

Table 1: Extract of film descriptions data (D_1)

*The mark scale is given as an indication.

	title	hasActor	rDate	director	lang
i_{12}	Ocean's 11	J. Roberts; B. Pitt;	2001	S. Soderbergh	
i_{22}	Ocean's 12	J. Roberts; B. Pitt	2004	S. Soderbergh; P.Greengrass	
i_{32}	Ocean's 13	B. Pitt; G. Clooney	2007	S. Soderbergh; P.Greengrass	
i_{52}	Bourne Identity		2002	P. Greengrass	en
i_{62}	Ocean's twelve	J. Roberts; B. Pitt; G. Clooney	2004		

Table 2: Extract of film descriptions data (D_2)

Question 3 (3 pts). Using the L2R method and considering the axiom $PFI(hasActor, director)$ of the class Film what would be the `owl:sameAs` links that can be obtained between the instances of D_1 and D_2?

Question 4 (2 pts). If you apply the property sharing rule of the sameAs predicate:

$$sameAs(x, y) \wedge p(x, z) \rightarrow p(y, z),$$

what would be the new property values that can be inferred?

2 Part 2: Ontology Alignment (6pts)

- *Question 5 (1.5 pt).* Give three kinds of heterogeneity in ontologies that can be faced when dealing with ontology alignment.
- *Question 6 (2 pts).* Given the ontology alignment problem shown in Figure 1, (i) explain the different inputs and (ii) give two examples of relations that can be used to represent mappings in A'.

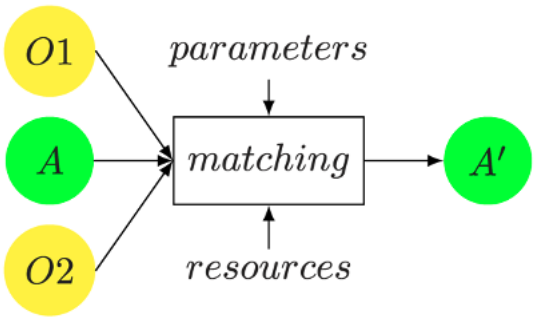

Figure 1: Ontology alignment problem

- Let us consider two ontologies $O1$ and $O2$ of Figure 2. In table 3, we give the set of identity links between instances of these two ontologies.
  - *Question 7 (1.5 pt).* If we apply an instance-based ontology alignment what would be the ontology mappings between the classes of these two ontology that can be found?

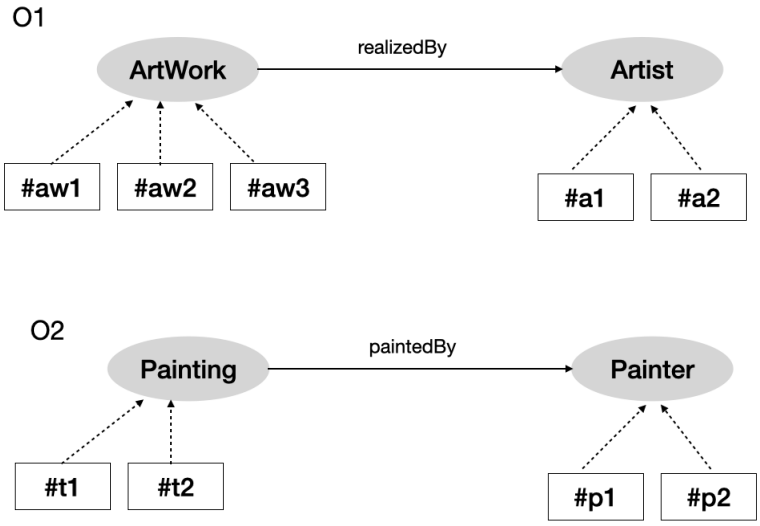

Figure 2: Two ontologies O1 and O2

SameAs(#aw1, #t1)	SameAs(#aw2, #t2)
SameAs(#a1, #p1)	SameAs(#a2, #p2)

Table 3: Identity links of the instances of O_1 and O_2

– *Question 8 (1 pt)* In the same setting, what would be the ontology mappings between the properties of these two ontologies that you may suggest?

3 Part 3: Link invalidation (5pt)

- *Question 9 (1.5 pts).* Give three reasons that may lead to incorrect sameAs links.
- *Question 10 (1.5 pts).* Give the four properties that define the semantics of `sameAs` predicate.
- *Question 11 (2 pts).* According to the recent literature studies, cite three different kinds of approaches that can be used to detect erroneous identity links.