Practical Session 2: Web of Data - Data Linkning

Pablo Mollá Chárlez

Contents

1 Part 1: Data Linking

Let us consider the two following data sources S_1 and S_2 presented in Figure 1.

Source S1 :
Library(L11), Book(b11), Book(b12) City(c11)
name(L11, *"François Mitterrand"*);
contains(L11,b11);
locatedAt(L11,c11);
cityName(c11,"Paris");
address(L11,"*Quai François Mauriac, 75706 Paris*");
title(b11, *"Les misérables"*);
publishedBy(b11, *"A. Lacroix"*);
title(m11, *"New York Times"*);
fondationDate(m11,*1862*); title(b12, *"Madame de bovary"*);

Source S2 :
Library(L21), Book(b21),Book(b22), City(c11),
name(L21, *"Bibliothèque Nationale de France"*);
locatedAt(L21,c21);
contains(L21, b21);
contains(L21, b22);
cityName(c21, *"Ville de paris"*);
title(b21, *"Madame bovary"*);
title(b22, *"Les misérables'*);
publishedBy(b22, *"Albert Lacroix"*);
publishedBy(b21, *"Michel Lévy frères"*);
adresseMusee(L21,"*15 Rue Emile Durkheim, 75013 Paris*");
title(m21, *"NYT"*);
fondationDate(m21,*1862*);

Figure 1: Example of Data Sources: Books & Libraries

Consider the ontology of Figure 2 as representing the vocabulary used to describe the facts of S_1 and S_2. Additionally, consider also the following axioms as declared in the ontology of Figure 2:

Property Type	**Properties**
Functional Properties	a_1: PF(title) a_2: PF(publishedBy) a_3: PF(cityName) a_4: PF(editionDate) a_5: PF(foundationDate) a_6: PF(name)
Inverse Functional Properties	a_7: PFI(contains) a_8: PFI(name) a_9: PFI(title, foundationDate) a_{10}: PFI(title, publishedBy)

Table 1: Functional and Inverse Functional Properties

Let us consider as well the following **synVals** facts:

synVals('Madame bovary', 'Madame de bovary'), synVals('Les misérables', 'Les misérables'),
synVals('A. Lacroix', 'Albert Lacroix'), synVals('New York Times', 'NYT'),
synVals(1862, 1862)

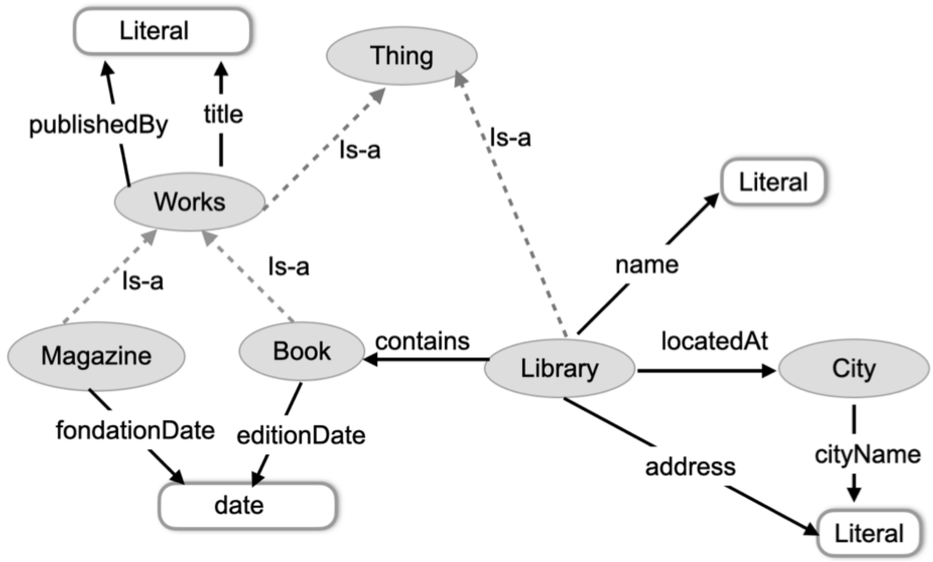

Figure 2: Library Ontology

- **Question 1.** If you consider the domain and range definition of the property **fondationDate**, what would be the **rdf:type** of $m11$ and $m21$?
- **Question 2.** Give the logical rules that can be generated for the axioms: $a1$, $a3$, $a7$ and $a10$.
- **Question 3.** If you consider the **UNA** as stated in both S_1 and S_2, the logical rules that can be generated for all the axioms $a1$ to $a10$, the RDF facts in S_1 and S_2 as well as the above **synVals** facts, what would be the identity links (sameAs) that you can obtain if you apply L2R Method (the logical method for data linking)?
- **Question 4.** What are the additional identity links that one can infer if we consider an additional axiom **a11:PF(locatedAt)**?

1.1 Answers

- **Question 1.** If you consider the domain and range definition of the property **fondationDate**, what would be the **rdf:type** of $m11$ and $m21$?

  As inferred from the Figure 1, the instances **fondateDate(m11,1862)** and **fondationDate(m21,1862)** are instanced by the property **fondationDate** who's range is a Literal (literals represent a value such as a string, number, or date) of type date. Besides, from the figure 2, the domain of **fondationDate** are instances of the class Magazine. We can conclude that, the **rdf:type** of $m11$ and $m21$ is Magazine.

- **Question 2.** Give the logical rules that can be generated for the axioms: $a1$, $a3$, $a7$ and $a10$.

  Let's recall the definitions of Functional Property (PF) and Inverse Functional Property (PFI):

  1. Functional Property (PF) A property **P** is functional, if for an individual **X**, there is only one value for **P**. Mathematically, it's defined as:

  $$\forall X, Y, Z \quad P(X,Y) \land P(X,Z) \implies Y = Z$$

  2. Inverse Functional Property (PFI) A property **P** is inverser functinoal if the inverse of **P** is functional. Mathematically, it's defined as:

  $$\forall X, Y, Z \quad P(X,Y) \land P(Z,Y) \implies X = Z$$

We will provide all logical rules that can be generated from the axioms a_1 to a_{10}:

- a_1:**PF(title)** $\longleftrightarrow$ sameAs(X, Y) $\wedge$ title(X, Z) $\wedge$ title(Y, W) $\Longrightarrow$ synVals(W,Z)
  The meaning of $title(X, Z)$ is "The title of X is Z".
- a_2:**PF(publishedBy)** $\longleftrightarrow$ sameAs(X, Y) $\wedge$ publishedBy(X, Z) $\wedge$ publishedBy(Y, W) $\Longrightarrow$ synVals(W,Z)
- a_3:**PF(cityName)** $\longleftrightarrow$ sameAs(X, Y) $\wedge$ cityName(X, Z) $\wedge$ cityName(Y, W) $\Longrightarrow$ synVals(W,Z)
- a_4:**PF(editionDate)** $\longleftrightarrow$ sameAs(X, Y) $\wedge$ editionDate(X, Z) $\wedge$ editionDate(Y, W) $\Longrightarrow$ synVals(W,Z)
- a_5:**PF(foundationDate)** $\longleftrightarrow$ sameAs(X, Y) $\wedge$ foundationDate(X, Z) $\wedge$ foundationDate(Y, W) $\Longrightarrow$ synVals(W,Z)
- a_6:**PF(name)** $\longleftrightarrow$ sameAs(X, Y) $\wedge$ name(X, Z) $\wedge$ name(Y, W) $\Longrightarrow$ synVals(W,Z)
  Now, when it comes to the logical rule of inverse functional properties, it becomes a bit tricky.
- a_7:**PFI(contains)** $\longleftrightarrow$ sameAs(Z, W) $\wedge$ iscontained(Z, X) $\wedge$ iscontained(W, Y) $\Longrightarrow$ sameAs(X,Y)
  In this logical rule, $iscontained(Z, X)$ is understood as Z is contained in X, which is the inverse property of $contains(X, Z)$ where X contains Z.
- a_8:**PFI(name)** $\longleftrightarrow$ synVals(N_1, N_2) $\wedge$ isnamed(L_{11}, N_1) $\wedge$ isnamed(L_{12}, N_2) $\Longrightarrow$ sameAs(L_{11}, L_{12})
- a_9:**PFI(title, foundationDate)** $\longleftrightarrow$ synVals(T_1, T_2) $\wedge$ synVals(D_1, D_2) $\wedge$ title(M_{11}, T_1) $\wedge$ title(M_{12}, T_2) $\wedge$ foundationDate(M_{11}, D_1) $\wedge$ foundationDate(M_{12}, D_2) $\Longrightarrow$ sameAs(M_{11}, M_{12})
- a_{10}:**PFI(title, publishedBy)** $\longleftrightarrow$ synVals(T_1, T_2) $\wedge$ synVals(D_1, D_2) $\wedge$ title(B_{11}, T_1) $\wedge$ title(B_{12}, T_2) $\wedge$ publishedBy(B_{11}, D_1) $\wedge$ publishedBy(B_{12}, B_2) $\Longrightarrow$ sameAs(B_{11}, B_{12})

We need to take into account that when infering the logical rules, if we are considering classes then we will be using in both sides of the implication sameAs, and on the other hand, if we are using one class and an instance, then depends on the implication, i.e., if the implication is related to instances then we use synVals, and if is related to classes we use sameAs.

- **Question 3.** If you consider the **UNA** as stated in both S_1 and S_2, the logical rules that can be generated for all the axioms $a1$ to $a10$, the RDF facts in S_1 and S_2 as well as the above **synVals** facts, what would be the identity links (sameAs) that you can obtain if you apply L2R Method (the logical method for data linking)?

  In order to answer to this question, we need to have in mind the previous logical rules so we can identify clearly which instances can be used. For instance, we can infer the following **sameAs** and **synVals** identity links:

$$\bigstar \left\{ \begin{array}{l} \text{synVals("A. Lacroix", "Albert Lacroix")} \wedge \text{synVals("Les misérables", "Les misérables")} \wedge \\ \text{title(b11, "Les misérables")} \wedge \text{title(b22, "Les misérables")} \wedge \text{publishedBy(b11, "A. Lacroix")} \wedge \\ \text{publishedBy(b22, "Albert Lacroix")} \underbrace{\Longrightarrow}_{a_{10}} sameAs(b11, b22) \end{array} \right.$$

$$\left\{ \begin{array}{l} \text{synVals(1862, 1862)} \wedge \text{synVals("New York Times", "NYT")} \wedge \text{title(m11, "New York Times")} \wedge \\ \text{title(m21, "NYT")} \wedge \text{fondationDate(m11, 1862)} \wedge \text{publishedBy(m21, 1862)} \underbrace{\Longrightarrow}_{a_9} sameAs(m11, m21) \end{array} \right.$$

We remind that, **contains(L11, b11)** $\longleftrightarrow$ **iscontained(b11, L11)**.

$$\blacksquare \left\{ \text{iscontained(b11, L11)} \wedge \text{iscontained(b22, L21)} \wedge \underbrace{\text{sameAs(b11, b22)}}_{\bigstar} \underbrace{\Longrightarrow}_{a_7} sameAs(L21, L11) \right.$$

Finally, the last possible identity link, which is a $synVals$ instead of a $sameAs$ instance:

$$\left\{ \begin{array}{l} \text{name(L11, "François Mittérrand")} \wedge \text{name(L21, "Bibliothèque Nationale de France")} \wedge \\ \underbrace{\text{sameAs(L21, L11)}}_{\blacksquare} \underbrace{\Longrightarrow}_{a_6} \text{synVals('François Mittérrand', 'Bibliothèque Nationale de France')} \end{array} \right.$$

- **Question 4.** What are the additional identity links that one can infer if we consider an additional axiom **a11:PF(locatedAt)**?

  Considering the additional axiom **a11:PF(locatedAt)** as a functional property allows us to obtain the following instance:

$$\begin{cases} \text{LocatedAt(L11, c11)} \ \wedge \ \text{LocatedAt(L21, c21)} \ \wedge \ \underbrace{\text{sameAs(L21, L11)}}_{\blacksquare} \\ \Longrightarrow \text{LocatedAt(L11, c21)} \wedge \text{LocatedAt(L11, c21)} \underbrace{\Longrightarrow}_{PF} sameAs(c11, c21) \end{cases}$$

2 Part 2: Combination of Ontology Alignment and Data Linking

Let's consider O_1, O_2 and O_3 three populated ontologies in the academic domain that we show in Figure 3. In figure 4, we give the set of identity links between instances of these three ontologies.

- **Question 5.** If you apply the transitivity property on the identity links that you obtain what would be the additional sameAs links that one can get (you can consider that symmetry property already applied)?

- **Question 6.** If we apply an instance-based ontology alignment what would be the ontology mappings that one can obtain between the classes of O_1, O_2 and O_3 ?

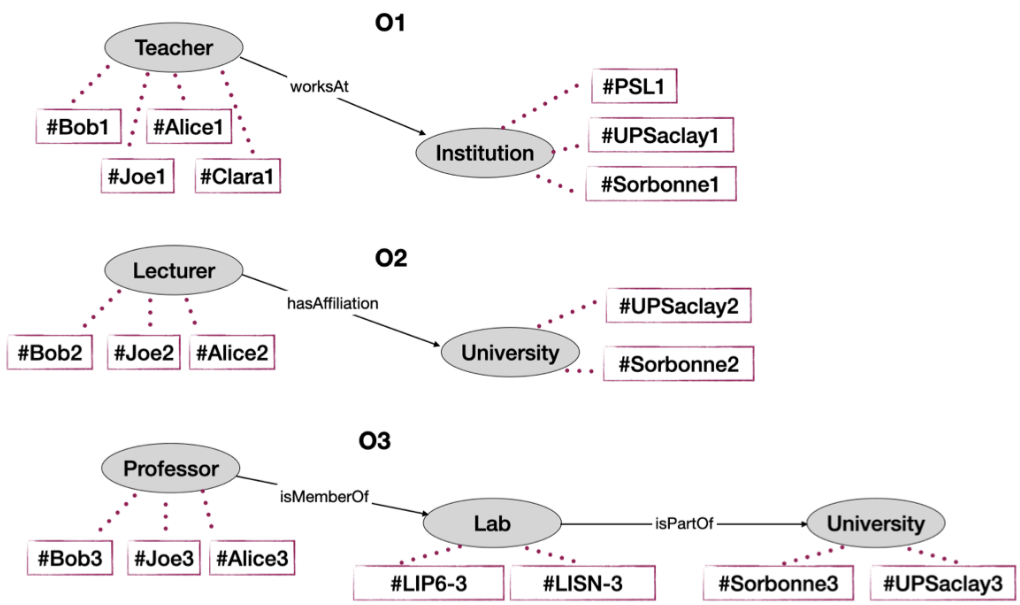

Figure 3: University Ontologies O_1, O_2 and O_3

SameAs(#Bob1, #Bob2)	SameAs(#Bob2, #Bob3)
SameAs(#Alice1, #Alice2)	SameAs(#Alice2, #Alice3)
SameAs(#Joe2, #Joe3)	SameAs(#Joe3, #Joe1)
SameAs(#Clara1, #Clara2)	SameAs(#Clara3, #Clara1)
SameAs(#UPSaclay1, #UPSaclay2)	SameAs(#UPSaclay2, #UPSaclay3)
SameAs(#Sorbonne1, #Sorbonne2)	SameAs(#Sorbonne2, #Sorbonne3)

Figure 4: Identity links of the instances of O_1, O_2 and O_3

2.1 Answers

- **Question 5.** If you apply the transitivity property on the identity links that you obtain what would be the additional **sameAs** links that one can get (you can consider that symmetry property already applied)?

  The additional **sameAs** links that can be obtained after applying the transitivity property as well as using the symmetry property are:

  - $sameAs(\#Bob1, \#Bob2) \wedge sameAs(\#Bob2, \#Bob3) \implies sameAs(\#Bob1, \#Bob3)$
  - $sameAs(\#Alice1, \#Alice2) \wedge sameAs(\#Alice2, \#Alice3) \implies sameAs(\#Alice1, \#Alice3)$
  - $sameAs(\#Joe2, \#Joe3) \wedge sameAs(\#Joe3, \#Joe1) \implies sameAs(\#Joe2, \#Joe1)$
  - $sameAs(\#Clara1, \#Clara2) \wedge sameAs(\#Clara3, \#Clara1)$
    $\longleftrightarrow sameAs(\#Clara2, \#Clara1) \wedge sameAs(\#Clara1, \#Clara3) \implies sameAs(\#Clara2, \#Clara3)$
  - $sameAs(\#UPSaclay1, \#UPSaclay2) \wedge sameAs(\#UPSaclay2, \#UPSaclay3)$
    $\implies sameAs(\#UPSaclay1, \#UPSaclay3)$
  - $sameAs(\#Sorbonne1, \#Sorbonne2) \wedge sameAs(\#Sorbonne2, \#Sorbonne3)$
    $\implies sameAs(\#Sorbonne1, \#Sorbonne3)$

- **Question 6.** If we apply an instance-based ontology alignment what would be the ontology mappings that one can obtain between the classes of O_1, O_2 and O_3 ?

  Let's remind a bit of theory, particularly about the possible instance-based relations:

  - **Equivalence Mapping** ($\equiv$): There is a bijection between both ontologies.
  - **Subsumption Mapping** ($\subseteq$ or $\supseteq$): One ontology O_1 is contained within another ontology O_2, i.e., all the mapping instances from O_1 are contained within O_2, but O_2 might contain even more.
  - **Disjoint Mapping** ($\perp$): Both ontologies are explicitly different.

  From the set of **sameAs** links, we can separate the comparison of the three ontologies to analyze the instance-based alignment and then the property-based alignment:

  1. $\boxed{O_1 \ \& \ O_2}$
     All O_2**:Lecturer** instances have a corresponding instance in O_1**:Teacher** and the instance **#Clara1** in O_1**:Teacher** is the only instance which doesn't have a corresponding instance in O_2**:Lecturer**, then we can deduce that:

     O_2**:Lecturer** **SubClassOf** O_1**:Teacher**

     Analogously, we have another inclusion between the set of instances of O_2**:University** and O_1**:Institution**, meaning that:

     O_2**:University** **SubClassOf** O_1**:Institution**

  2. $\boxed{O_1 \ \& \ O_3}$
     For these two ontologies, we observe that:

     O_3**:Professor** **SubClassOf** O_1**:Teacher**

     As well as:

     O_3**:University** **SubClassOf** O_1**:Institution**

  3. $\boxed{O_2 \ \& \ O_3}$
     In this comparison, we can observe the existence of an equivalence, because we have a bijection between the sets of the instances of O_2**:Lecturer** and O_3**:Professor**, then:

     O_2**:Lecturer** **equivalentClass** O_3**:Professor**

Similarly, we have another bijection as follows:

$$O_2\textbf{:University} \quad \textbf{equivalentClass} \quad O_3\textbf{:University}$$

Finally, in terms of the property-based alignment, we observe that:

$$O_2\textbf{:hasAffiliation} \quad \textbf{subPropertyOf} \quad O_1\textbf{:WorksAt}$$

$$O_3\textbf{:isMemberOf/}O_3\textbf{:isPartOf} \quad \textbf{subPropertyOf} \quad O_1\textbf{:WorksAt}$$

$$O_3\textbf{:isMemberOf/}O_3\textbf{:isPartOf} \quad \textbf{equivalentProperty} \quad O_2\textbf{:hasAffiliation}$$