

TIW2 - TD

Alignement d'ontologies

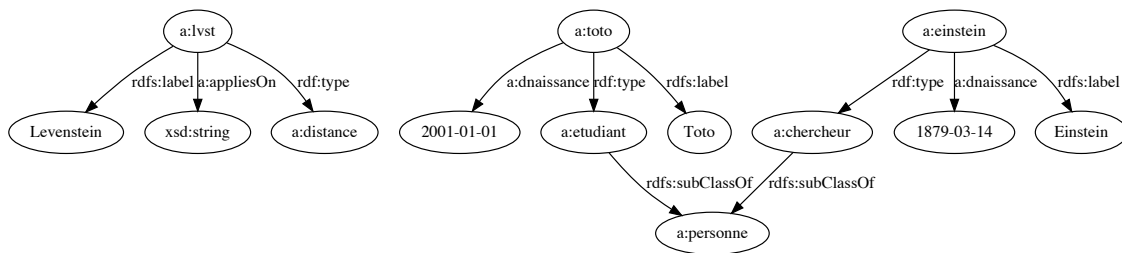
Exercice 1:

On considère les deux ontologies suivantes

Ontologie A

```
@prefix a: <http://example.com/univ/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema> .

a:lvst rdfs:label "Levenstein";
      a:appliesOn xsd:string;
      rdf:type a:distance.
a:toto rdfs:label "Toto";
      a:dnaissance "2001-01-01";
      rdf:type a:etudiant.
a:einstein rdfs:label "Einstein";
          a:dnaissance "1879-03-14";
          rdf:type a:chercheur.
a:etudiant rdfs:subClassOf a:personne.
a:chercheur rdfs:subClassOf a:personne.
```



Ontologie B

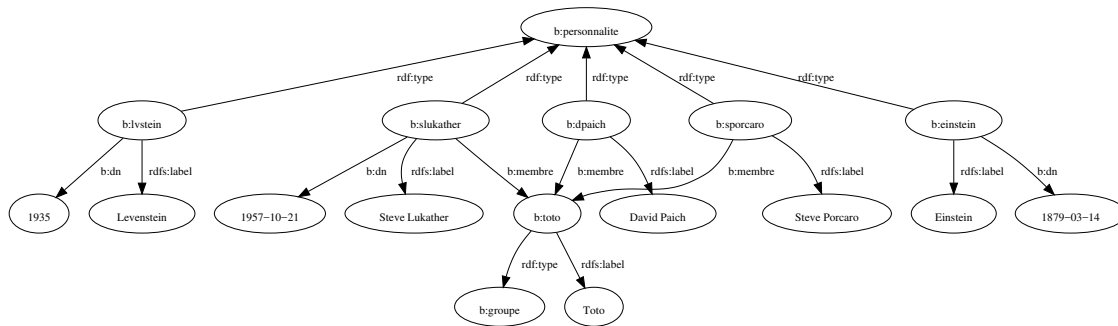
```
@prefix b: <http://personnes.com/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema> .

b:lvstein rdfs:label "Levenstein";
          b:dn "1935";
          rdf:type b:personnalite.
b:einstein rdfs:label "Einstein";
          b:dn "1879-03-14";
          rdf:type b:personnalite.
b:toto rdfs:label "Toto";
      rdf:type b:groupe.
```

```

b:slukather rdfs:label "Steve Lukather";
            b:membre b:toto;
            b:dn "1957-10-21";
            rdf:type b:personnalite.
b:dpaich rdfs:label "David Paich";
         b:membre b:toto;
         rdf:type b:personnalite.
b:sporcaro rdfs:label "Steve Porcaro";
           b:membre b:toto;
           rdf:type b:personnalite.

```



1. On rappelle la formule de similarité entre deux chaînes de caractères basée sur les n -gram :

$$\bar{\sigma}(s, t) = \frac{|n\text{-gram}(s) \cap n\text{-gram}(t)|}{\min(|s|, |t|) - n + 1}$$

avec $n\text{-gram}(s) = \{s' \in \text{strings}(s) \mid |s'| = n\}$

Calculer la similarité $n = 3$ pour les couples de labels suivants : ("Levenstein", "Levenstein"), ("Levenstein", "Einstein"), ("Levenstein", "Steve Lukather"), ("Levenstein", "Steve Porcaro"), ("Toto", "Toto"), ("Einstein", "Levenstein") et ("Einstein", "Einstein").

Pour les autres couples, elle vaut 0

2. On fait correspondre les instances dont la propriété `rdfs:label` est similaire à plus de 0,8.
3. Reconstituer les correspondances entre classes (en indiquant une inclusion \subseteq , l'égalité $=$ ou le partage d'instances \cap) sur la base de leur extension. Ces correspondances vous semblent-elles justes? D'où vient le problème?
4. Proposer une correspondance entre propriétés ($=$ ou \subseteq) en vous appuyant sur votre connaissance du domaine. En s'appuyant sur cette correspondance, proposer une moyenne pondérée pour comparer les instances de façon à faire disparaître le problème rencontré ci-dessus. Vérifier en reconstituant à nouveau la correspondance entre classes via les instances.
5. Proposer des correspondances entre classes en vous basant sur votre connaissance du domaine. Ces correspondances permettent-elles de résoudre le problème apparu en 2.? Proposer un (des) axiome(s) dans les ontologies qui permettrait de détecter le problème (par exemple en rendant incohérente l'ontologie résultant de l'alignement).