

Présentation

Projet de web sémantique

UFR Sciences et Techniques - Nantes
M1 ALMA - 2017/2018

Le Bars Yannis, Travers Clément, Fortin Guillaume

L'objectif de ce projet est de transformer des données ouvertes de l'Enseignement supérieur, de la Recherche et de l'Innovation en données sémantiques et de lier ces données sémantiques au cloud de "Linked Data : Connect Distributed Data across the Web"

- **Etape 1** : choisir un ensemble de données ouvertes à « sémantifier » et proposer deux requêtes SPARQL « intéressantes » réalisables sur ces données.
- **Etape 2** : lier nos données à des données des autres groupes et proposer une requête réalisable sur au moins 2 ensemble de données.

Etape 1 : choisir un ensemble de données ouvertes à « sémantifier »

Description du sujet

Sujet : Appels à projets Horizon 2020 - Projets retenus et participants identifiés des appels du Programme-Cadre de Recherche et d'Innovation (PCRI)

Nombre d'enregistrements : 19.101

Nombre d'attributs : 22

Date de création : 1 juin 2017

Le dataset regroupe les projets ayant commencé entre le 1 janvier 2014 et le 1 juin 2016 (dans les faits cette date a été étendu au 2 décembre 2017) et devant se terminer aux alentours de 2020 (entre 2014 et 2023 dans les faits).

Chaque projet a un ou plusieurs partenaires venant d'un ou plusieurs pays, une durée, un montant, un résumé, etc.

Lien : <https://data.enseignementsup-recherche.gouv.fr/explore/dataset/appels-a-projets-horizon-2020-projets-retenus-et-participants-identifies/information/>

Etape 1 : choisir un ensemble de données ouvertes à « sémantifier »

Méthode de réalisation

Outil utilisé: Tarql

On crée un code permettant de recevoir les données en entrée sous la forme d'un fichier csv, et qui renvoie en sortie les données sémantiques en rdf.

```
?URI rdf:type doap:Project;  
      frapo:hasCode ?code_du_projet;  
      doap:name ?titre;  
      dbp:shortName ?acronyme;  
      ?URI7 ?URI6;  
      time:hasBeginning ?date_de_debut;  
      time:hasEnd ?date_de_fin;  
      doap:description ?resume;  
      time:Duration ?duree;  
      dbp:number ?montant;  
      foaf:topic ?theme;  
      ?URI8 ?code_rcn;  
      doap:homepage ?lien_cordis;  
      dcat:keyword ?mots_cles;  
      org:linkedTo ?URI2.
```

```
?URI6 rdf:type owl:Class;  
      rdfs:label ?appel_a_projet;  
      frapo:hasCode ?code_d_appel_a_projet.
```

```
?URI7 rdf:type owl:DatatypeProperty.
```

Notre première URI identifie un projet. Nous nous sommes aidés d'exemples trouvés sur les internets. ?code_du_projet, ?titre, ?acronyme, ... sont les propriétés du projet. Pour les cas spéciaux comme ?appel_a_projet et ?code_d_appel_a_projet, nous avons défini notre propre vocabulaires spécifiques car nous n'avons pas trouvé de vocabulaires correspondant. Pour le reste, nous avons utilisé du vocabulaire préexistants : dpb, dpo, owl, doap, foaf, rdf ...

L'URI7 s'appelle "Appel_Projet", c'est une "DatatypeProperty". L'URI6 contient l'appel du projet et le code de l'appel.

```
?URI2 rdf:type dbo:Organisation;  
      ?URI5 ?identifiant_de_partenaire;  
      rdfs:label ?libelle_de_partenaire;  
      dbo:type ?URI3;  
      dbo:locationCountry ?URI4.
```

```
?URI3 rdf:type owl:Thing;  
      rdfs:label ?type_de_partenaire;  
      rdfs:label ?code_du_type_de_partenaire.
```

```
?URI4 rdf:type dbo:PopulatedPlace;  
      rdfs:label ?pays_du_partenaire;  
      rdfs:label ?code_pays.
```

```
?URI5 rdf:type owl:DatatypeProperty.  
?URI8 rdf:type owl:DatatypeProperty;  
      rdfs:range xsd:nonNegativeInteger.
```

Notre deuxième URI est de type Organisation, elle est liée aux projets avec le vocabulaire org:linkedTo, c'est pour représenter les partenaires qui sont liés aux projets.

Notre troisième URI permet l'identification des types de partenaire des projets.

Notre quatrième permet l'identification des pays des partenaires du projet. Nous avons pas lié les pays à des URI déjà existantes, car ?pays_du_partenaire et ?code_pays sont fait de valeurs multiples.

URI5 identifie les participants par leur numéro siren ou autre selon leur nature.

URI8 est un vocabulaire pour le code rcn et c'est un entier non négatif.

Etape 1 : proposer deux requêtes SPARQL « intéressantes »

Première requête

Requête pour retourner les 100 projets qui ont reçus le plus de subvention par rapport à la durée du projet.

```
SELECT ?titre ?duree ?montant
FROM <donnees.rdf>
WHERE {
    ?x doap:name    ?titre;
        time:Duration ?duree;
        dbp:number   ?montant.
    BIND((?montant/?duree) AS ?rapport)
}
ORDER BY DESC(?rapport)
LIMIT 100
```

Etape 1 : proposer deux requêtes SPARQL « intéressantes »

Deuxième requête

Deuxième requête

```
SELECT ?titre
FROM <csv.csv>
WHERE {
  ?x doap:name ?titre
  BIND ((year (?date__de__fin) - year (?date__de__debut))*12 + month
  (?date__de__fin) - month (?date__de__debut) AS ?mois__effectif)
  BIND (?mois__effectif - ?duree AS ?difference )
  ORDER BY ?difference
```


Etape 2 : lier nos données

Nous avons choisi de lier notre dataset avec celui du groupe de Brohan Romain et Gascoin-Fontaine Samy, car le sujet de leur dataset est très similaire au nôtre, il porte également sur les participants à un programme-cadre de recherche.

sujet du groupe : Les projets retenus et participants identifiés lors du 7ème appel à projet du Programme-Cadre de Recherche et de Développement Technologique (PCRDT).

Etape 2 : proposer une requête

Requête pour retourner les thèmes des projets avec les pays associés.

```
SELECT ?theme ?pays_du_partenaire
FROM NAMED <http://notre\_dataset>
FROM NAMED <http://leur\_dataset>
WHERE {
    ?x foaf:topic ?theme;
        org:linkedTO ?partenaire.
    ?partenaire dbo:locationCountry ?pays.
    ?pays rdfs:label ?pays_du_partenaire
}
```

FIN