

Web Sémantique et Social : Interrogation de données RDF

TP à rendre en binômes le 18/03 à minuit sur Moodle.

Le rendu devra être composé de deux parties :

i) Le présent TP

ii) Le TP du 8/03

Les deux parties sont à rendre au même moment (18/03) et dans un seul document pdf (synthétique).

1) Interrogation d'un flux RSS exporté en RDF

Connectez vous à l'endpoint SPARQL générique à l'adresse :

<http://sparql.org/sparql.html>. Spécifiez l'option `output="Text"`.¹

Le document RDF, à l'adresse <http://www.w3.org/2001/sw/SW-FAQ-feed.rdf>, correspond à un *flux RSS*² permettant de générer la page <http://www.w3.org/RDF/FAQ>.

Vu en RDF, un *flux RSS* est une instance de la classe `rss:channel`. Il contient des articles, instances de la classe `rss:item`.

Ci-dessous une première requête pour tester le fonctionnement de l'endpoint :

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rss: <http://purl.org/rss/1.0/>
PREFIX swq: <http://www.w3.org/2001/sw/SW-FAQ#>

SELECT DISTINCT ?varClass
FROM <http://www.w3.org/2001/sw/SW-FAQ-feed.rdf>
WHERE { ?s rdf:type ?varClass }
```

Exprimez les requêtes suivantes à l'aide de SPARQL.

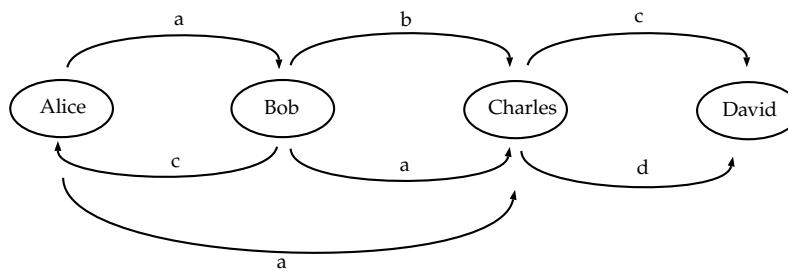
1. Quels sont les titres des articles (`rss:item`) publiés dans le flux RSS ?
2. Quel est le titre (`rss:title`) du flux (`rss:channel`) RSS ?
3. Donnez les onze premiers articles du flux RSS par ordre chronologique.
4. Donnez le deuxième dernier article du flux RSS (ordre chronologique) ainsi que l'avant dernier (UNION).
5. Quelles sont les paires d'articles qui ont été publiées à la même date ?
6. Quels sont les articles publiés le 2007-04-12 ?
(Suggestion : Format pour la date dans la requête `"20070412T00:00+00:00"^^xsd:string`)
7. Donnez la liste des auteurs (sans répétition) des articles dans ce flux RSS ?
8. Dites si le flux RSS utilise la propriété `"title"` définie par l'espace de noms Dublin Core (dublincore.org/).
9. Dites si le flux RSS utilise la propriété `"subject"` définie par l'espace de noms Dublin Core.
10. Donnez la liste des dates de publication des articles dans le flux RSS.

Dans votre rendu, il vous suffira de reporter les requêtes qui correspondent aux interrogations, sans leurs résultats.

1. En alternative à l'endpoint vous pouvez utiliser Jena. Voir la classe `QueryRemoteRDF.java` dans le Moodle

2. <https://fr.wikipedia.org/wiki/RSS>

Figure 1



2) SPARQL Property-paths

Donnez le résultat de l'évaluation des requêtes suivantes dans le graphe en Figure 1.

1. `SELECT ?y WHERE { Alice (!b/!b/!d)* ?y . }`
2. `SELECT ?y WHERE { Alice (a|b|^c)* ?y . }`
3. `SELECT ?y WHERE { Alice ((a/c?/d)|(!a*)) ?y . }`
4. Indiquez trois chemins permettant de rejoindre Alice à partir de Charles.
5. Donnez une requête ASK qui vérifie l'existence d'un chemin (orienté) simple passant par les quatre noeuds du graphe.

3) Exploration d'un Endpoint SPARQL. Exemple de DBPedia.

Un problème classique dans le Web Sémantique est celui d'obtenir une description de l'organisation des données qu'on s'apprête à interroger.

Une solution pragmatique consiste à exécuter une série de requêtes simples et génériques sur les données afin d'obtenir un aperçu sur les descriptions utilisées.

L'objectif de cet exercice est de récupérer une description de la structure d'un fragment de DBPedia.

Connectez vous à l'endpoint SPARQL de DBPedia (<http://dbpedia.org/sparql>) et testez-en le bon fonctionnement avec la requête suivante.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT DISTINCT ?varInstance
WHERE { ?varInstance rdf:type owl:Class }
LIMIT 100

```

Voici ensuite une trace à suivre pour l'exploration de DBPedia.

1. Sélectionnez la liste des propriétés de l'ontologie.
2. Choisissez deux classes de l'ontologie et donnez la liste de leurs sous-classes.
3. Comptez le nombre d'instances de chaque sous-classe, et triez-les en ordre croissant.
4. Trouvez les propriétés employées pour décrire les instances de la classe choisie.
5. Trouvez ensuite les paires de propriétés les plus utilisées conjointement.
6. Vérifiez si ces propriétés ont des sous-propriétés (requête 4).

Bases de Données Liées :

7. Cherchez des références vers des entrepôts de données RDF "externes" (autres que DBPedia).
8. Combien de liens externes retrouvez vous (requête 7) ?

Dans votre rendu, reportez les requêtes correspondantes aux interrogations, ainsi que les premières (max 10) lignes des résultats.