

BD2

TP6 : SPARQL

Yoann Couillec

8 avril 2015

1 Introduction

Il s'agit d'apprendre à manipuler le langage SPARQL.

SPARQL est un langage de requête pour le Web sémantique. Sa syntaxe ressemble à SQL afin d'être accessible à tous, mais il faut comprendre que le modèle de données sous-jacent est différent du modèle relationnel. Il s'agit d'une organisation des données sous forme de graphe.

Un graphe est un ensemble de noeuds reliés entre eux par des arêtes (dirigés ou non). Dans le cas du Web sémantique, les noeuds sont identifiés par des Uniform Resource Identifier (URI) similaire à un Uniform Resource Locator (URL).

Par exemple, Les informations de DBpedia concernant Barack Obama sont accessible via son URI :

`http://dbpedia.org/resource/Barack_Obama`

Lorsque vous accédez à cette ressource dans un navigateur vous êtes redirigé vers une page mettant en forme les informations. L'URL de cette page est :

`http://dbpedia.org/page/Barack_Obama`

Il est tout à fait possible d'obtenir les informations sous format JSON ou XML.

2 Questions

2.1 Question 1

Aller dans un navigateur et accéder à la ressource :

`http://dbpedia.org/resource/Barack_Obama`

Constater la redirection.

2.2 Question 2

Faire de même en utilisant `telnet`.

```
> telnet dbpedia.org 80
GET /resource/Barack_Obama HTTP/1.1
Host: dbpedia.org
```

Vous obtenez la réponse :

```
HTTP/1.1 303 See Other
Date: Wed, 08 Apr 2015 08:58:55 GMT
Content-Type: text/html; charset=UTF-8
Content-Length: 0
Connection: keep-alive
Server: Virtuoso/07.20.3213 (Linux) x86_64-redhat-linux-gnu VDB
Location: http://dbpedia.org/page/Barack_Obama
Expires: Wed, 15 Apr 2015 08:58:55 GMT
Cache-Control: max-age=604800
```

Le code 303 signifie “See other”, c’est le code de redirection. L’entête “Location” contient l’URL à aller voir.

Se rendre à cette URL avec `telnet`. La réponse est un fichier HTML. La sauvegarder dans le fichier “barack.html” en utilisant la redirection du flux de sortie “>”. Afficher dans un navigateur le fichier sauvegardé. Les CSS ne sont pas téléchargés donc l’affichage est laid mais les informations sont bien présentes.

Vous constatez qu’un navigateur ne fait rien de plus que ce que vous venez de faire avec `telnet`.

Bonus (pendant les vacances) Écrire votre programme `telnet` en langage C. Il suffit d’utiliser les sockets.

2.3 Question 3

Accéder à la ressource Barack Obama en demandant le format JSON :

```
GET /resource/Barack_Obama HTTP/1.1
Host:dbpedia.org
Accept:application/json
```

La réponse est :

```
HTTP/1.1 303 See Other
Date: Wed, 08 Apr 2015 09:10:08 GMT
Content-Type: application/json; qs=0.6
Content-Length: 0
Connection: keep-alive
Server: Virtuoso/07.20.3213 (Linux) x86_64-redhat-linux-gnu VDB
```

TCN: choice
Vary: negotiate,accept
Alternates: ...
Link: ...
Location: http://dbpedia.org/data/Barack_Obama.json
Expires: Wed, 15 Apr 2015 09:10:08 GMT
Cache-Control: max-age=604800

Il s'agit d'une redirection. Accéder à la nouvelle URL et sauvegarder la réponse dans le fichier "barack.json". L'ouvrir dans Chrome. Vous constatez que Chrome détecte que vous lui demandez d'ouvrir un fichier JSON et propose un affichage adapté.

2.4 Question 4

Récupérer les données de Barack Obama au format Turtle et sauvegarder dans le fichier "barack.ttl".

2.5 Question 5

Ouvrir un navigateur et se rendre à l'URL :

<http://dbpedia.org/sparql>

Vous obtenez une interface permettant de requêter le graphe de DBpedia en SPARQL.

La requête par défaut est :

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

Les variables commencent par "?". Le nombre de résultats maximum est 100. La clause **where** contient les informations cruciales de la requête. En fait, la clause **where** contient une succession de motifs de graphe. Dans le cas de la requête par défaut [] est le noeud de départ, **a** est l'arête et **?Concept** est le noeud d'arrivée. On peut préciser plusieurs motifs en les séparant par un point :

```
select ?x where { ?x a foaf:Person . ?x foaf:name "Aaron Lines"@en }
```

a signifie que l'on cherche les noeuds **?x** dont le type est une personne. Ensuite, on précise le nom de la personne. Et on récupère son URI.

Écrire et exécuter une requête permettant de trouver Michael Jackson. La réponse que l'on veut est :

http://dbpedia.org/resource/Michael_Jackson

2.6 Question 6

Écrire et exécuter une requête permettant de donner la date de naissance de Michael Jackson.

2.7 Question 7

Écrire et exécuter une requête permettant de donner les coordonnées de la ville de naissance de Michael Jackson. Les propriétés à trouver sont `geo:lat` et `geo:long`. Vous remarquez qu'il faut effectuer une "jointure" entre 2 noeuds, le noeud "Michael Jackson" et le noeud de sa ville de naissance.

2.8 Question 8

Trouver les villes de plus de 10 millions d'habitants.

2.9 Questions 9

Donner le nom des oeuvres de Leonardo Da Vinci.

2.10 Bonus (pendant la retraite)

Écrire un programme donnant l'arbre des types. La propriété est :

```
rdfs:subClassOf
```