

SPARQL

Cours « Recherche d'Information et Graphe de données »

Nicolas Delestre

Plan...

- 1 Publication de données RDF
- 2 SPARQL Query Language
- 3 SPARQL Update
- 4 Conclusion

Point de vue : document → base de données 1 / 2

Point de vue document

- Si URI dérérérençable, on obtient un document le décrivant en utilisant le protocole HTTP
- Problèmes :
 - comment connaître cet URI ?
 - comment obtenir uniquement les informations (triplés qui nous interesse)
 - comment faire pour obtenir des informations sur un URI qui n'est pas dérérérençable ?
 - comment stocker ces informations lorsqu'il y a des milliers voire des millions de triplets RDF ?
 - comment mettre à jour ces informations ?

Point de vue : document → base de données 2 / 2

Point de vue base de données

- Stocker les triplés RDF dans des bases de données spécialisées (appelées *triple stores* ou *RDF stores*)
- Interroger ces bases de données à l'aide d'un langage de requête adapté : SPARQL
- Continuer à utiliser le Web : les requêtes et réponses utilisent le protocole HTTP

Réponse du W3C

SPARQL

Documentations du W3C

- *SPARQL Query Language* : présentation du langage de requête SPARQL
- *SPARQL Update* : présentation du langage de mise à jour des données
- *SPAQRL Service Description* : présentation du vocabulaire de description d'un service SPARQL
- *SPARQL Protocol* : présentation du protocole de communication entre client et serveur SPARQL
- *SPARQL Graph Store HTTP Protocol* : présentation du protocole de communication entre client et serveur pour les opérations de gestion de graphes
- *SPARQL Query Results XML Format/JSON/CSV Format* : présentation des formats de résultats de requêtes SPARQL
- *SPARQL Entailment Regimes* : présentation des différentes méthodes qui permettent de trouver les triplés réponses

SPARQL

- Syntaxe ressemblant à SQL
- Principe de fonctionnement : recherche de sous-graphes dans un graphe (*graph matching*)

SPARQL par l'exemple

- Tous les exemples suivants sont à tester sur Wikidata :
 - Tests en ligne : <https://query.wikidata.org/>
 - Tests depuis un programme : <https://query.wikidata.org/sparql>
- Issus de « Du Web des documents au Web Sémantique », ils ont pour objectifs l'obtention d'informations sur Michel Bussi

Michel Bussi sur Wikidata

Michel Bussi - Wikidata - Chromium

Michel Bussi - Wikidata x

https://www.wikidata.org/wiki/Q3309133

Item [Discussion](#) [Read](#) [View history](#)

Wikidata

[Main page](#)
[Community portal](#)
[Project chat](#)
[Create a new item](#)
[Item by title](#)
[Recent changes](#)
[Random item](#)
[Query Service](#)
[Nearby](#)
[Help](#)
[Donate](#)

Tools
[What links here](#)
[Related changes](#)

Michel Bussi (Q3309133)

French writer, University professeur

[In more languages](#) Configure

Language	Label	Description	Also known as
English	Michel Bussi	French writer, University professeur	
French	Michel Bussi	écrivain français, professeur des universités	
Spanish	Michel Bussi	No description defined	
German	Michel Bussi	französischer Schriftsteller, Politologe, Geograph und Hochschullehrer	

[All entered languages](#)

Les préfixes

Les préfixes les plus courants

```
PREFIX wd: <http://www.wikidata.org/entity/>  
PREFIX wdt: <http://www.wikidata.org/prop/direct/>  
PREFIX p: <http://www.wikidata.org/prop/>  
PREFIX ps: <http://www.wikidata.org/prop/statement/>  
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
```

Exemples d'URI de Wikidata

```
wd:Q3309133 Michel Bussi  
  wdt:P50 auteur (a été écrit par)  
  wdt:P1476 titre
```

Exemples de programme

En Python

```
#!/usr/bin/python3
# -*- coding: utf-8 -*-
from SPARQLWrapper import SPARQLWrapper, JSON
import sys

query = """
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
"""

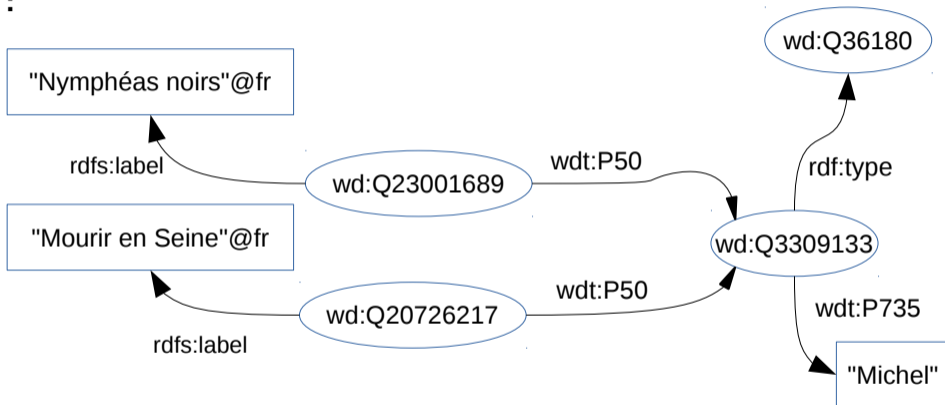
with open(sys.argv[1]) as f:
    for line in f:
        query = query + line

sparql = SPARQLWrapper("https://query.wikidata.org/sparql")
sparql.setQuery(query)
sparql.setReturnFormat(JSON)
results = sparql.query().convert()

for result in results["results"]["bindings"]:
    st = ""
    for key in result.keys():
        st = st + key + " : " + result[key][u'value'] + " , "
    print(st[:-3])
```

Principe d'une requête SPARQL

Base :



Requête :

Les URI des livres écrits par Michel Bussi

Requête

```
SELECT ?uri WHERE {  
  ?uri wdt:P50 wd:Q3309133. # a été écrit par M. Bussi  
}
```

Résultats

```
uri : http://www.wikidata.org/entity/Q2981427  
uri : http://www.wikidata.org/entity/Q16681699  
uri : http://www.wikidata.org/entity/Q20726217  
uri : http://www.wikidata.org/entity/Q23001462  
uri : http://www.wikidata.org/entity/Q23001689  
uri : http://www.wikidata.org/entity/Q23001767  
uri : http://www.wikidata.org/entity/Q23653447  
uri : http://www.wikidata.org/entity/Q23653619  
uri : http://www.wikidata.org/entity/Q23653620  
uri : http://www.wikidata.org/entity/Q23653624  
uri : http://www.wikidata.org/entity/Q23653647  
uri : http://www.wikidata.org/entity/Q23653915  
uri : http://www.wikidata.org/entity/Q24082328
```

Les titres des livres écrits par Michel Bussi

Requête

```
SELECT ?titre WHERE {  
  ?uri wdt:P50 wd:Q3309133. # a été écrit par M. Bussi  
  ?uri wdt:P1476 ?titre. # a pour titre  
}
```

Résultats

```
titre : Code Lupin  
titre : Un avion sans elle  
titre : Death on the Seine  
titre : Mourir sur Seine  
titre : Omaha crimes  
titre : Nymphéas noirs  
titre : Ne lâche pas ma main  
titre : Sang famille  
titre : N'oublier jamais  
titre : Gravé dans le sable  
titre : Maman a tord  
titre : Éléments de géographie électorale à travers l'exemple de la France de l'Ouest  
titre : Pour une nouvelle géographie du politique : Territoire, démocratie, élections.  
titre : Le temps est assassin
```



Les titres, en français, des livres écrits par Michel Bussi

Requête

```
SELECT ?titre WHERE {  
  ?uri wdt:P50 wd:Q3309133. # a été écrit par M. Bussi  
  ?uri wdt:P1476 ?titre. # a pour titre  
  FILTER (LANGMATCHES(LANG(?titre),"FR"))  
}
```

Résultats

```
titre : Code Lupin  
titre : Un avion sans elle  
titre : Mourir sur Seine  
titre : Omaha crimes  
titre : Nymphéas noirs  
titre : Ne lâche pas ma main  
titre : Sang famille  
titre : N'oublier jamais  
titre : Gravé dans le sable  
titre : Maman a tord  
titre : Éléments de géographie électorale à travers l'exemple de la France de l'Ouest  
titre : Pour une nouvelle géographie du politique : Territoire, démocratie, élections.  
titre : Le temps est assassin
```



Les titres, en français, des romans écrits par Michel Bussi

Requête

```
SELECT ?titre WHERE {  
  ?uri wdt:P50 wd:Q3309133. # a été écrit par M. Bussi  
  ?uri wdt:P1476 ?titre. # a pour titre  
  ?uri wdt:P136 wd:Q8261. # est de genre roman  
  FILTER (LANGMATCHES(LANG(?titre),"FR"))  
}
```

Résultats

titre : Un avion sans elle

Les titres, en français, des romans (ou romans policiers) écrits par Michel Bussi

Requête

```
SELECT ?titre WHERE {
  ?uri wdt:P50 wd:Q3309133. # a été écrit par M. Bussi
  ?uri wdt:P1476 ?titre. # a pour titre
  VALUES (?type_livre) {(wd:Q208505) (wd:Q8261)}
  ?uri wdt:P136 ?type_livre. # est de genre
  FILTER (LANGMATCHES(LANG(?titre),"FR"))
}
```

Résultats

```
titre : Code Lupin
titre : Un avion sans elle
titre : Mourir sur Seine
titre : Omaha crimes
titre : Nymphéas noirs
titre : Ne lâche pas ma main
titre : Sang famille
titre : N'oublier jamais
titre : Gravé dans le sable
titre : Maman a tord
titre : Le temps est assassin
```

Les titres et éditeurs des livres écrits par Michel Bussi

Requête

```
SELECT ?titre ?nom_editeur WHERE {
  ?uri_livre wdt:P50 wd:Q3309133. # a été écrit par M. Bussi
  ?uri_livre wdt:P1476 ?titre. # a pour titre
  FILTER (LANGMATCHES(LANG(?titre),"FR"))
  ?uri_livre wdt:P123 ?uri_editeur. # est édité par
  ?uri_editeur wdt:P1448 ?nom_editeur. # a pour nom
}
```

Résultats

```
titre : Un avion sans elle , nom_editeur : Presses de la cité
titre : Nymphéas noirs , nom_editeur : Presses de la cité
titre : Ne lâche pas ma main , nom_editeur : Presses de la cité
titre : N'oublier jamais , nom_editeur : Presses de la cité
```

Les titres et optionnellement l'éditeur des livres écrits par Michel Bussi

Requête

```

SELECT ?titre ?nom_editeur WHERE {
  ?uri_livre wdt:P50 wd:Q3309133. # a été écrit par M. Bussi
  ?uri_livre wdt:P1476 ?titre. # a pour titre
  FILTER (langmatches(lang(?titre),"FR"))
  OPTIONAL {
    ?uri_livre wdt:P123 ?uri_editeur. # éditeur
    ?uri_editeur wdt:P1448 ?nom_editeur. # nom officiel
  }
}

```

Résultats

```

titre : Un avion sans elle , nom_editeur : Presses de la cité
titre : Nymphéas noirs , nom_editeur : Presses de la cité
titre : Ne lâche pas ma main , nom_editeur : Presses de la cité
titre : N'oublier jamais , nom_editeur : Presses de la cité
titre : Code Lupin
titre : Mourir sur Seine
titre : Omaha crimes
titre : Sang famille
titre : Gravé dans le sable
titre : Maman a tort
titre : Éléments de géographie électorale à travers l'exemple de la France de l'Ouest
titre : Pour une nouvelle géographie du politique : Territoire, démocratie, élections.
titre : Le temps est assassin

```

Et ce n'est pas fini !

SPARQL Query Language

- Possibilité d'enchaîner plusieurs propriétés (opérateur /)
- Possibilité de répéter plusieurs fois une propriété (opérateur +)
- Possibilité d'utiliser des fonctions dans le SELECT : MIN, MAX, COUNT, ...
- Possibilité d'ordonner les résultats : ORDER BY
- Possibilité de grouper des résultats (GROUP BY) et d'agréger des résultats (HAVING) avec différentes fonctions d'agrégation (MAX, AVG, SUM, ...)
- Possibilité de limiter le nombre de résultats LIMIT (à utiliser sur des grandes base) ou de demander les résultats à partir d'un certain rang (OFFSET)
- Possibilité d'utiliser des variables : AS
- Possibilité de rechercher dans plusieurs bases à la fois : SERVICE

Quelques instructions en plus du SELECT 1 / 2

ASK : répond par vrai ou faux

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
ASK {
  wd:Q3309133 wdt:P50 ?auteur .
}
```

DESCRIBE : décrit un URI

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
DESCRIBE wd:Q3309133
```

Attention

- Le résultat d'une requête DESCRIBE dépend de l'implémentation du serveur SPARQL ciblé



Quelques instructions en plus du SELECT 2 / 2

CONSTRUCT : construit un graphe à partir d'une requête

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
CONSTRUCT {
  wd:Q3309133 wdt:P50 ?auteur .
}
WHERE {
  wd:Q3309133 wdt:P50 ?auteur .
}
```

SPARQL Update

- SPARQL Update est un langage de mise à jour de données RDF
- Il permet d'ajouter, de supprimer ou de modifier des triplés RDF
- Encore faut-il avoir les droits pour le faire sur la base de données ciblée (souvent restreint à l'administrateur de la base de données)

SPARQL Update par l'exemple 1 / 2

Ajout

```
PREFIX ex: <http://example.org/>

INSERT DATA {
  ex:alice ex:city "Paris" .
}
```

Suppression

```
PREFIX ex: <http://example.org/>

DELETE DATA {
  ex:alice ex:age 24 .
}
```

SPARQL Update par l'exemple 2 / 2

Suppression sous condition

```
PREFIX ex: <http://example.org/>
```

```
DELETE {  
  ex:alice ex:age ?age .  
}  
WHERE {  
  ex:alice ex:age ?age .  
  FILTER(?age < 18)  
}
```

Mise à jour = suppression + insertion

```
PREFIX ex: <http://example.org/>
```

```
DELETE {  
  ex:alice ex:age 24 .  
}  
INSERT {  
  ex:alice ex:age 25 .  
}  
WHERE {  
  ex:alice ex:age 24 .  
}
```

Conclusion

Quelques points d'entrée SPARQL

- DBpédia : <https://dbpedia.org/sparql>
- Wikidata : <https://query.wikidata.org/>
- DBnary (données wikitionnaire) : <http://kaiko.getalp.org/sparql>
- DBLP : <http://dblp.13s.de/d2r/snorql>
- INSEE : <http://rdf.insee.fr/sparql>
- LinkedGeoData (données OpenStreetMap) : <http://linkedgeo.org/sparql>
- MeSH : <https://id.nlm.nih.gov/mesh/query>
- EMBL-EBI (European Bioinformatics Institute) : <http://www.ebi.ac.uk/rdf/services/sparql>
- ...

Catalogue de données

- Datahub : <https://old.datahub.io/>

Attention

- À la disponibilité des points d'entrée SPARQL
- Au temps de réponse des points d'entrée SPARQL ⇒ utiliser l'instruction LIMIT
- À la qualité des données (incomplet, erroné, ...)