

Le web sémantique

SPARQL

Triplestores

- On peut représenter une BD sous forme relationnelle
 - Langage de requêtes : SQL
- On peut aussi utiliser un ensemble de triplets
 - Triplestore : entrepôt de triplets
 - Base de données NoSQL
 - Schéma de BD : types (classes) et propriétés
 - Langage de requêtes : SPARQL
 - Moteur de SGBD ↔ SPARQL Endpoint (ex : Jena)
 - Volume parfois colossal (wikidata)

SPARQL

- Langage de requêtes adapté aux triplets
- « Sparkle » : étincelle
- « SPARQL Protocol and RDF Query Language »
- Recommandation W3C (2008)
- Version 1.1 : 2013

Exemple

- Quel est le concept nommé en Français : « Chat » ?
- Dans le TP N-Triples :

? <http://www.w3.org/2000/01/rdf-schema#label> "Chat"@fr .

? rdfs:label "Chat"@fr .

Exemple

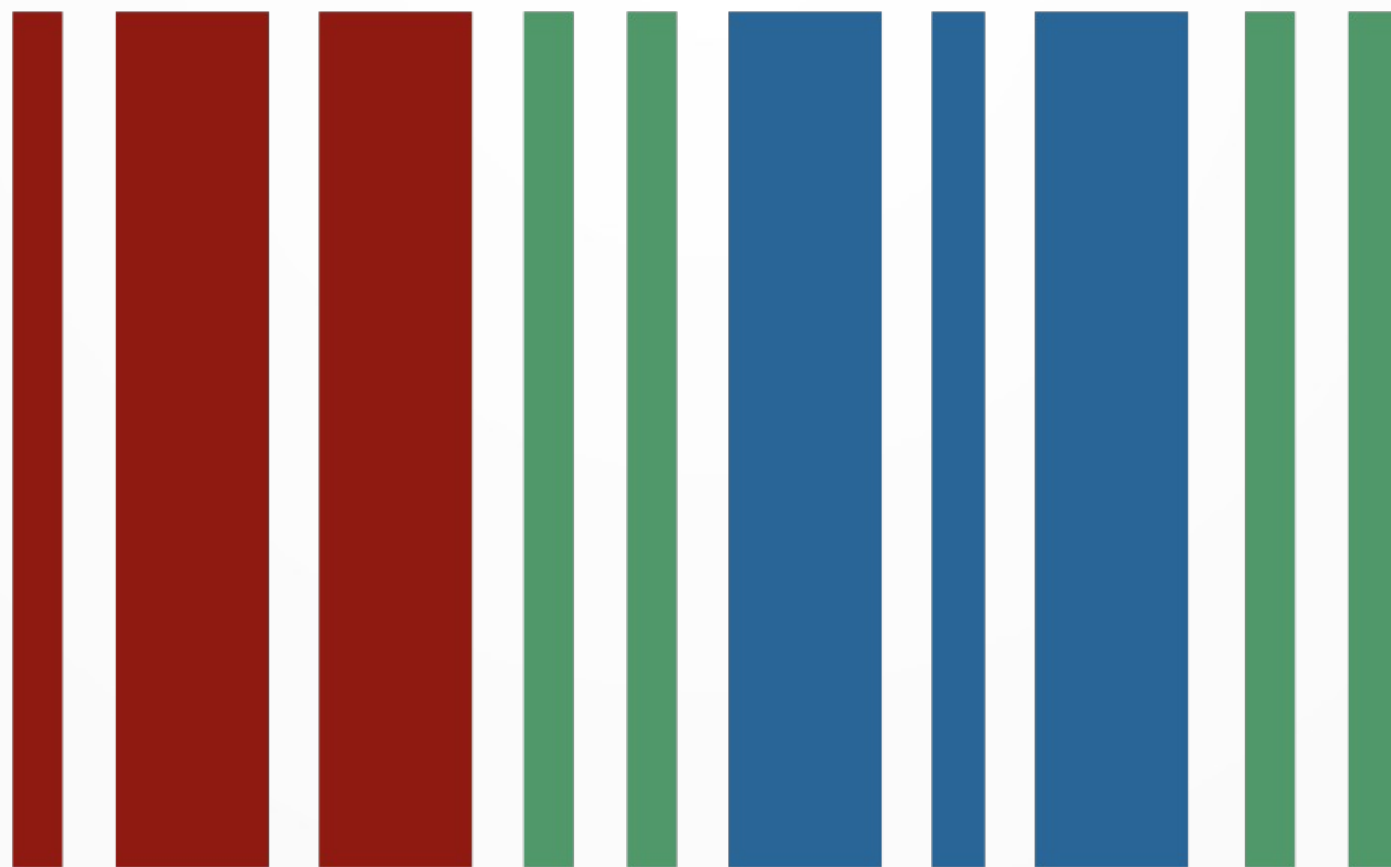
- Quel est le concept nommé en Français : « Chat » ?
- En SPARQL :

```
SELECT ?s WHERE {  
    ?s rdfs:label "Chat"@fr .  
}
```

Résultat

- En SPARQL, comme en SQL, on renvoie des listes de tuples
 - Autant de lignes que de résultats
 - Autant de colonnes que de variables
- On donne une liste de contraintes
 - Triplets qui doivent être vrais dans la base
 - Variables à instancier

Wikidata



WIKIDATA

Wikidata

- <https://www.wikidata.org>
- Version « web sémantique » de wikipédia et autres projets wikimédia
- Plusieurs centaines de millions de triplets
- Peut être exportée vers N-Triples, RDF/XML ou autres
- Peut être interrogée en ligne via SPARQL

Exemple

- Est-ce qu'il existe une valeur s telle que le triplet $?s$ `rdfs:label "Chat"@fr` est vrai ?

```
SELECT ?s WHERE {  
    ?s rdfs:label "Chat"@fr .  
}
```

Chat : il y en a plein !

- Chat (Q1838423) : village in Kyrgyzstan
- Chat (Q405417) : Wikimedia disambiguation page
- Felis (Q1642433) : former constellation
- Chat (Q14863672) : protein-coding gene in the species *Mus musculus*
- Etc.
- Il n'y a même pas le bon !

Exemple

- À la conquête du chat domestique

```
SELECT ?s WHERE {  
    ?s rdfs:label "chat domestique"@fr .  
}
```

→ wd:Q146

Exemple

- De quoi le chat domestique est-il une sous-classe ?

```
SELECT ?s ?classe WHERE {  
    ?s rdfs:label "chat domestique"@fr .  
    ?s wdt:P279 ?classe .  
}
```

wdt:P279 = équivalent de rdfs:subClassOf dans WD

Exemple

- De quoi le chat domestique est-il une sous-classe ?

```
SELECT ?s ?classe WHERE {  
    ?s rdfs:label "chat domestique"@fr ;  
    wdt:P279 ?classe .  
}
```

→ ?s ?classe

wd:Q146 wd:Q25265

wd:Q146 wd:Q39201

wd:Q146 wd:Q57814795

Exemple

- De quoi le chat domestique est-il une sous-classe ?
- Récupérer les noms des classes plutôt que les classes elles-mêmes

```
SELECT ?labelClasse WHERE {  
    ?s rdfs:label "chat domestique"@fr ;  
    wdt:P279 ?classe .  
    ?classe rdfs:label ?labelClasse .  
}
```

Exemple

- De quoi le chat domestique est-il une sous-classe ?
- Récupérer les noms des classes plutôt que les classes elles-mêmes

```
SELECT ?labelClasse WHERE {  
    ?s rdfs:label "chat domestique"@fr ;  
    wdt:P279 [  
        rdfs:label ?labelClasse  
    ] .  
}
```

Oups !

- Il faudrait pouvoir filtrer selon la langue !

- Чэтыу
- Felidae
- የፍመት አስተኔ
- Felidae
- Catt
- سنوریات
- السنوریات
- Felidae
- krapoleem
- Pişiklər
- پیشیکلر
- каціныя
- каціныя
- Коткови
- ...

Exemple

- Quel est le nom de l'auteur du roman intitulé *Swag* ?

```
SELECT ?authorName WHERE {  
  ?swag rdfs:label "Swag"@en ;  
  wdt:P31 wd:Q7725634 ; # wdt:P31 = rdf:type dans WD  
  wdt:P50 [  
    rdfs:label ?authorName  
  ] .  
}
```

Tous les romans de l'auteur de *Swag*

```
SELECT ?book WHERE {  
    ?swag rdfs:label "Swag"@en ;  
        wdt:P31 wd:Q7725634 ;  
        wdt:P50 ?author .  
    ?book wdt:P50 ?author .  
}
```

Filtrer les résultats

- Tous les romans dont l'auteur s'appelle Elmore Leonard

```
SELECT ?book WHERE {  
  ?book wdt:P31 wd:Q7725634;  
  wdt:P50 [  
    rdfs:label "Elmore Leonard"@en  
  ].  
}
```

Filtrer les résultats

- Tous les romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais.

```
SELECT ?book WHERE {  
  ?book wdt:P31 wd:Q7725634;  
  wdt:P50 [  
    rdfs:label "Elmore Leonard"@en  
  ] ;  
  rdfs:label ?title.  
  FILTER (?title != "Swag"@en)  
}
```

Résultat

- wd:Q4000827
- wd:Q4000827
- wd:Q5123382
- wd:Q5123382
- wd:Q5569989
- wd:Q5887854
- wd:Q5887854
- wd:Q5887854
- wd:Q5887854
- wd:Q6407902
- wd:Q6407902
- wd:Q6407902
- ...

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais.

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
  wdt:P50 [  
    rdfs:label "Elmore Leonard"@en  
  ] ;  
  rdfs:label ?title.  
  FILTER (?title != "Swag"@en)  
}
```

Résultat

- The Complete Western Stories
- Tutti i racconti western
- City Primeval: High Noon in Detroit
- City Primeval
- Glitz
- Hombre
- Hombre
- Hombre
- Hombre (nofel)
- Il Corvo
- Killshot
- Killshot
- 生死一擊
- ...

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais, et merci de ne me donner que les titres en Anglais !

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
    wdt:P50 [  
      rdfs:label "Elmore Leonard"@en  
    ] ;  
  rdfs:label ?title.  
  FILTER (?title != "Swag"@en)  
  FILTER (lang(?title) = "en")  
}
```


Résultat

- The Complete Western Stories
- City Primeval
- Glitz
- Hombre
- Killshot
- LaBrava
- Pagan Babies
- Pronto
- Riding the Rap
- Road Dogs
- The Big Bounce
- The Hunted
- ...

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais, et merci de ne me donner que les titres en Français !

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
    wdt:P50 [  
      rdfs:label "Elmore Leonard"@en  
    ] ;  
  rdfs:label ?title.  
  FILTER (?title != "Swag"@en)  
  FILTER (lang(?title) = "fr")  
}
```

Résultat

- Pronto
- Beyrouth-Miami
- Djibouti
- Tous les titres n'ont pas été renseignés en Français...
- On voudrait les titres en Anglais, mais aussi en Français quand ils sont disponibles

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard et dont le titre commence par « Out of »

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
    wdt:P50 [  
      rdfs:label "Elmore Leonard"@en  
    ] ;  
  rdfs:label ?title.  
  FILTER (regex(?title, "^Out of"))  
  FILTER (lang(?title) = "en")  
}
```

→ Out of sight

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard et dont le titre **ne** commence **pas** par « Out of »

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
    wdt:P50 [  
      rdfs:label "Elmore Leonard"@en  
    ] ;  
  rdfs:label ?title.  
  FILTER (!regex(?title, "^Out of"))  
  FILTER (lang(?title) = "en")  
}
```

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard et dont le titre **ne** commence **pas** par « Out of », peu importe la casse

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
  wdt:P50 [  
    rdfs:label "Elmore Leonard"@en  
  ] ;  
  rdfs:label ?title.  
  FILTER (!regex(?title, "^out of", "i")) # i : case insensitive  
  FILTER (lang(?title) = "en")  
}
```

Filtrer les résultats

- Tous les romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais (rappel)

```
SELECT ?book WHERE {  
  ?book wdt:P31 wd:Q7725634;  
  wdt:P50 [  
    rdfs:label "Elmore Leonard"@en  
  ] ;  
  rdfs:label ?title.  
  FILTER (?title != "Swag"@en)  
}
```

Filtrer les résultats : MINUS

- Tous les romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais

```
SELECT ?book WHERE {  
  ?book wdt:P31 wd:Q7725634;  
  wdt:P50 [  
    rdfs:label "Elmore Leonard"@en  
  ] ;  
  rdfs:label ?title.  
MINUS {?book rdfs:label "Swag"@en}  
}
```


MINUS

- Notation ensembliste
- On a deux ensembles de résultats
- On retire du premier les éléments du second

Filtrer les résultats

- Tous les titres des romans dont l'auteur s'appelle Elmore Leonard, sauf celui qui s'appelle « Swag » en anglais, et merci de ne me donner que les titres en Français !
- Rappel : On n'avait pas tout, on veut les titres en Français *quand ils sont disponibles*

```
SELECT ?title WHERE {  
  ?book wdt:P31 wd:Q7725634;  
        wdt:P50 [  
          rdfs:label "Elmore Leonard"@en  
        ] ;  
  rdfs:label ?title.  
  FILTER (?title != "Swag"@en)  
  FILTER (lang(?title) = "fr")  
}
```

OPTIONAL

```
SELECT ?title ?frenchTitle WHERE {  
  ?book wdt:P31 wd:Q7725634;  
        wdt:P50 [ rdfs:label "Elmore Leonard"@en ] ;  
        rdfs:label ?title .  
FILTER (?title != "Swag"@en)  
FILTER (lang(?title) = "en")  
OPTIONAL {  
  ?book rdfs:label ?frenchTitle .  
  FILTER (lang(?frenchTitle) = "fr")  
}  
}
```

OPTIONAL

- Contraintes optionnelles
- Si on trouve de quoi satisfaire ces contraintes, on retourne les résultats correspondants
- Si on trouve un résultat qui satisfait tout sauf les contraintes optionnelles, le résultat est retourné avec une valeur vide pour les variables optionnelles

Résultat

Pronto	Pronto
Riding the rap	Beyrouth-Miami
Djibouti	Djibouti
The complete western stories	
City Primeval	
...	

UNION

- Maintenant, on veut soit :
 - Le titre en Français s'il existe
 - Le titre en Anglais dans le cas contraire
- En d'autres termes, on veut :
 - Tous les titres en Français
 - *Et* tous les titres en Anglais des romans qui n'ont pas de titre Français

UNION

```
SELECT ?title WHERE {  
  {  
    ?book wdt:P50 [ rdfs:label "Elmore Leonard"@en ] ;  
    rdfs:label ?title.  
    FILTER (lang(?title) = "fr")  
  } UNION {  
    ?book wdt:P50 [ rdfs:label "Elmore Leonard"@en ] ;  
    rdfs:label ?title.  
    FILTER (lang(?title) = "en")  
  }  
}
```

Oups

- **Pronto**
- **Beyrouth-Miami**
- Djibouti
- The Complete Western Stories
- City Primeval
- Glitz
- Hombre
- Killshot
- LaBrava
- Pagan Babies
- **Pronto**
- **Riding the Rap**
- ...

UNION

```
SELECT ?title WHERE {  
  {  
    ?book wdt:P50 [ rdfs:label "Elmore Leonard"@en ] ;  
    rdfs:label ?title.  
    FILTER (lang(?title) = "fr")  
  } UNION {  
    ?book wdt:P50 [ rdfs:label "Elmore Leonard"@en ] ;  
    rdfs:label ?title.  
    FILTER (lang(?title) = "en")  
    MINUS {  
      ?book rdfs:label ?frenchTitle.  
      FILTER (lang(?frenchTitle) = "fr")  
    }  
  }  
}
```

Tri

- Liste des romans d'Elmore Leonard, par date de publication

```
SELECT ?book ?datePub WHERE {  
    ?book wdt:P31 wd:Q7725634;  
    wdt:P50 [ rdfs:label "Elmore Leonard"@en ] ;  
    wdt:P333 ?datePub .  
}  
ORDER BY ?datePub
```

Agrégats

- COUNT, SUM, MAX, AVG, GROUP BY, HAVING
- Prix moyen des livres, pour chaque auteur de la base

```
SELECT ?x (AVG(?y) AS ?avgY)
```

```
WHERE {
```

```
    ?book wdt:P50 ?x ;
```

```
    :price ?y .
```

```
}
```

```
GROUP BY ?x
```

Agrégats

- Nombre de livres, pour chaque auteur de la base

```
SELECT ?x (COUNT(?book) AS ?nbBooks)
WHERE {
    ?book wdt:P50 ?x .
}
GROUP BY ?x
```

Agrégats

- Nombre de livres, pour chaque auteur *prolifique* de la base

```
SELECT ?x (COUNT(?book) AS ?nbBooks)
WHERE {
    ?book wdt:P50 ?x .
}
GROUP BY ?x
HAVING (?nbBooks > 20)
```

Doublons

- Noms des auteurs ayant publié un roman depuis 2010

```
SELECT ?name
```

```
WHERE {
```

```
    ?book wdt:P50 [ rdfs:label ?name ] ;
```

```
        :yearPublished ?year .
```

```
    FILTER(?year >= 2010)
```

```
}
```

Doublons

- Noms des auteurs ayant publié un roman depuis 2010, sans doublon

```
SELECT DISTINCT ?name
```

```
WHERE {
```

```
    ?book wdt:P50 [ rdfs:label ?name ] ;
```

```
        :yearPublished ?year .
```

```
FILTER(?year >= 2010)
```

```
}
```

Limites

- Noms des auteurs ayant publié un roman depuis 2010, sans doublon, et j'en veux 10 au max

```
SELECT DISTINCT ?name
```

```
WHERE {
```

```
    ?book wdt:P50 [ rdfs:label ?name ] ;
```

```
        :yearPublished ?year .
```

```
    FILTER(?year >= 2010)
```

```
}
```

```
LIMIT 10
```


Beaucoup d'autres choses

- Ce qui est possible en SQL est possible en SPARQL
- Alternative valable aux bases de données relationnelles
- Mais BD relationnelles : technologies plus matures
- Triplestores pertinents
 - Quand on dispose naturellement de données sous forme de triplets
 - Quand le schéma de la BD évolue en permanence ou est très variable en fonction des données (pas de notion de table)