RDF Stores / DBpedia / SPARQL

Advanced Data Management (BTI7528)

Thomas Strub (strut1)

Stefan Guggisberg (guggs2)

Table of Contents

Introduction to SPARQL

SELECT, FROM, OPTIONAL, FILTER, UNION, ORDER BY

Source: https://www.w3.org/TR/rdf-sparql-query/

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?person ?name ?birthDate ?birthPlace
FROM <http://dbpedia.org>
WHERE {
 {
  ?person dbo:birthPlace dbr:Bern .
  OPTIONAL {
   ?person dbo:birthDate ?birthDate .
   FILTER(?birthDate < '2000-01-01'^^xsd:dateTime)
  }
 } UNION {
  ?person dbo:birthDate ?birthDate .
  FILTER(?birthDate = '1991-11-20'^^xsd:dateTime)
 }
 OPTIONAL { ?person rdfs:label ?name . }
 OPTIONAL { ?person foaf:name ?name . }
 OPTIONAL { ?person dbp:name ?name . }
 OPTIONAL { ?person dbp:birthName ?name . }
 OPTIONAL { ?person dbp:alternativeNames ?name . }
 OPTIONAL { ?person dbo:birthPlace ?birthPlace . }
 OPTIONAL { ?person dbo:birthDate ?birthDate . }
}
ORDER BY ?name
OFFSET 15 LIMIT 5
```

DBpedia Questions

- Part 1
  - Q1: Who was born in Bern before 2000-01-01 (with name and DOB)?
  - Q2: What are Junkerngasse's previous names, when was it built, how long is it and what is its postal code?
- Part 2
  - Q3: Who was both born in Bern and died there as well?

DBpedia Questions
Q1: Who was born in Bern before 2000-01-01 (with name and date of birth)?

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?name ?birthDate
WHERE {
 [] dbo:birthPlace dbr:Bern ;
  dbp:name ?name ;
  dbo:birthDate ?birthDate .
 FILTER(?birthDate < '2000-01-01'^^xsd:dateTime)
}
```

name	birthDate
"Kerim Seiler"@en	"1974-11-14"^^xsd:date
"Olivier Ladislas de Weck"@en	"1968-10-19"^^xsd:date
"de Weck, Olivier"@en	"1968-10-19"^^xsd:date
"Reinhard Gammenthaler"@en	"1953-05-09"^^xsd:date
"Gammenthaler, Reinhard"@en	"1953-05-09"^^xsd:date
"Michael von Graffenried"@en	"1957-05-07"^^xsd:date
"Robert Scheuermeier"@en	"1927-07-31"^^xsd:date
"Stephan Klossner"@en	"1981-05-30"^^xsd:date
...	...

DBpedia Questions

Q2: What are Junkerngasse's previous names, when was it built, how long is it and what is its postal code?

```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?formerName ?buildingStartDate ?lengthM
?lengthM ?postalCode
WHERE {
 dbr:Junkerngasse dbo:formerName ?formerName ;
  dbo:buildingStartDate ?buildingStartDate ;
  dbp:lengthM ?lengthM ;
  dbo:postalCode ?postalCode .
}
```

formerName	buildingStartDate	lengthM	postalCode
"Edle Gasse"@en	"1191"	300	"3011"
"Kilchgasse"@en	"1191"	300	"3011"

DBpedia Questions
Q3: Who was both born in Bern and died there as well?

```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?name
WHERE {
 [] dbo:birthPlace dbr:Bern ;
  dbo:deathPlace dbr:Bern ;
  dbp:name ?name .
}
```

```
-------------------------------
| name                        |
===============================
| "Lindt, Rodolphe"@en        |
| "Sigmund Freudenberger"@en  |
| "Freudenberger, Sigmund"@en |
| "Eduard Fischer"@en         |
| "Fischer, Eduard"@en        |
| "Burkhard, Balthasar"@en    |
| "Kurz, Rudolf Friedrich"@en |
| "Rudolph Friederich Kurz"@en|
| "Johann Rudolf Wyss"@en     |
| "Wyss, Johann Rudolf"@en    |
| ...                         |
-------------------------------
```

Pros / Cons of RDF Stores

Pros

- Arbitrary data
- Schemaless
- Relations / predicates
- Distributed data (e.g. FOAF)

Cons

- Data quality
- Weak specifications (e.g. DBpedia)
  - Unclear naming conventions (dbo vs dbp)
  - Undefined object types (resource vs literal)
  - Undefined literal types (date vs year vs string)

Lessons Learnt

- Relatively easy to dive in
- Getting details right can be complicated
- Very important to maintain a clean graph
- Even more important to document the rules

Discussion

- Does anyone have experience with RDF / SPARQL?
- Are there any questions or comments?

- Thank you for your patience.