Knowledge Graphs

Lecture 3 - Querying RDFS with SPARQL

3.2 Complex Queries with SPARQL

Prof. Dr. Harald Sack & Dr. Mehwish Alam
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
AIFB - Karlsruhe Institute of Technology
Autumn 2020

Knowledge Graphs
Lecture 3: Querying RDF(S) with SPARQL

The Semantic Web Technology Stack
(not a piece of cake...)
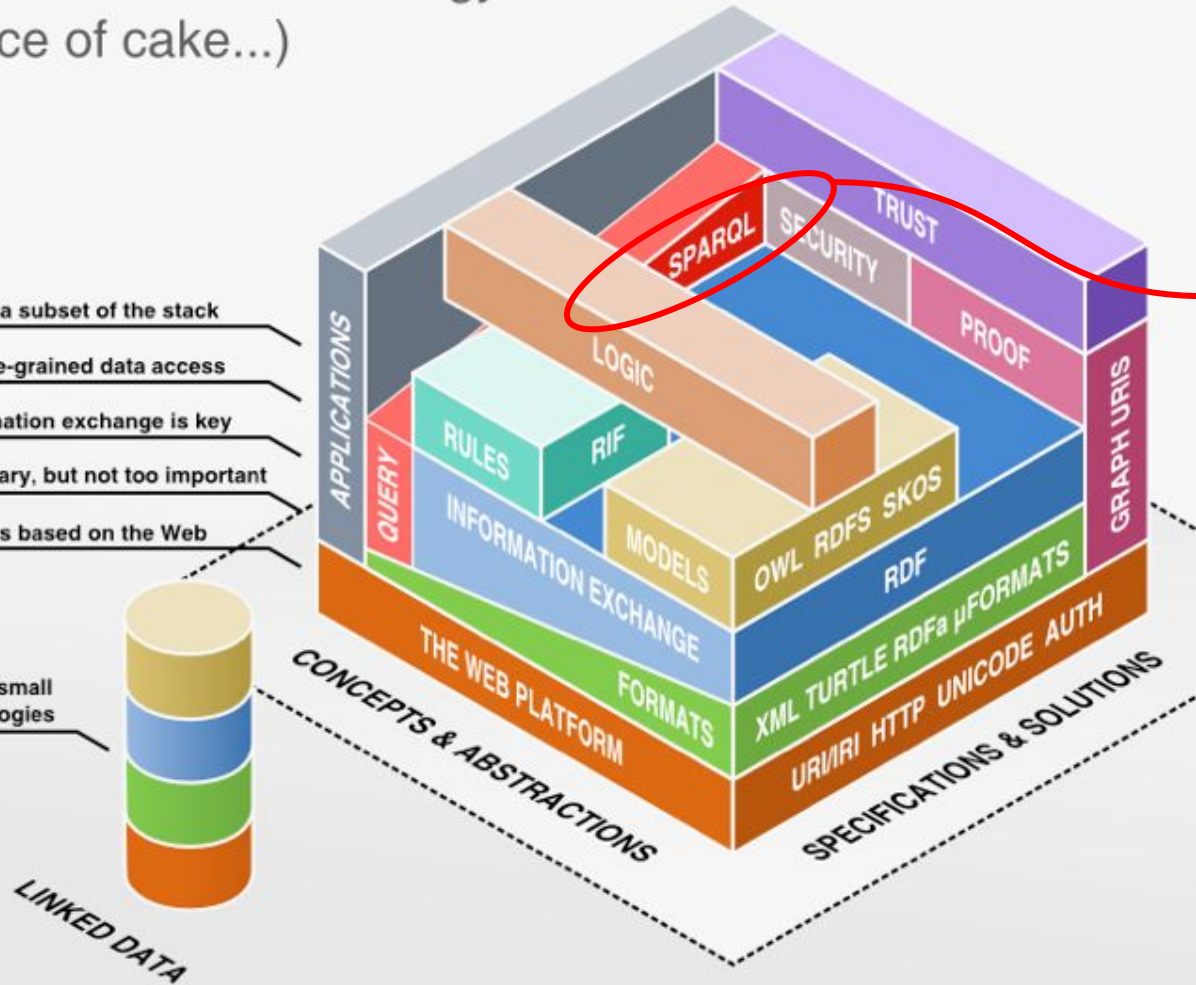

Most apps use only a subset of the stack

Querying allows fine-grained data access

Standardized information exchange is key

Formats are necessary, but not too important

The Semantic Web is based on the Web

Linked Data uses a small selection of technologies

USE **SPARQL** and QUERY ON

[1,2]

3

SPARQL Filter Constraints

```
PREFIX : <http://dbpedia.org/resource/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dct:  <http://purl.org/dc/terms/>
PREFIX dbc:  <http://dbpedia.org/resource/Category:>

SELECT ?author_name ?title
FROM <http://dbpedia.org/>
WHERE {
      ?author rdf:type dbo:Writer .
      ?author rdfs:label ?author_name
      FILTER (LANG(?author_name)="en").
      ?work dbo:author ?author .
          ?work rdfs:label ?title .
      FILTER (LANG(?title)="en")
      ?work dct:subject dbc:Environmental_fiction_books .
} LIMIT 100
```

- Example:
  Search for **authors** and their **books**, filter results for **English labels** and **Environmental Fiction Books** and **limit** the results to **the first 100**

query SPARQL endpoint

4

WIKIDATA Label Language Filtering

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX bd: <http://www.bigdata.com/rdf#>


SELECT ?authorLabel ?bookLabel ?date
WHERE {
    ?book wdt:P31 wd:Q47461344 .
    ?book wdt:P50 ?author .
    ?book wdt:P921 wd:Q7942 .
    ?book wdt:P577 ?date .
    SERVICE wikibase:label
    { bd:serviceParam wikibase:language "en" }
}
```

Instance of written work

author

main subject global warming

publication date

- Example:

Search for **authors and their books**, filter results for **English labels** and **Books on Global warming**

WIKIDATA

query SPARQL endpoint

5

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

WIKIDATA Label Language Filtering

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX bd: <http://www.bigdata.com/rdf#>

SELECT ?authorLabel ?bookLabel ?date
WHERE {
    ?book wdt:P31 wd:Q47461344 .
    ?book wdt:P50 ?author .
    ?book wdt:P921 wd:Q7942 .
    ?book wdt:P577 ?date .
    SERVICE wikibase:label
    { bd:serviceParam wikibase:language "en" }
}
```

wikidata specific label service

- Example:

  Search for **authors and their books**, filter results for **English labels** and **Books on Global warming**

WIKIDATA

query SPARQL endpoint

6

- Example:

  Search for **authors and their books**, filter results for **English labels** and

  Books on Global warming

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology
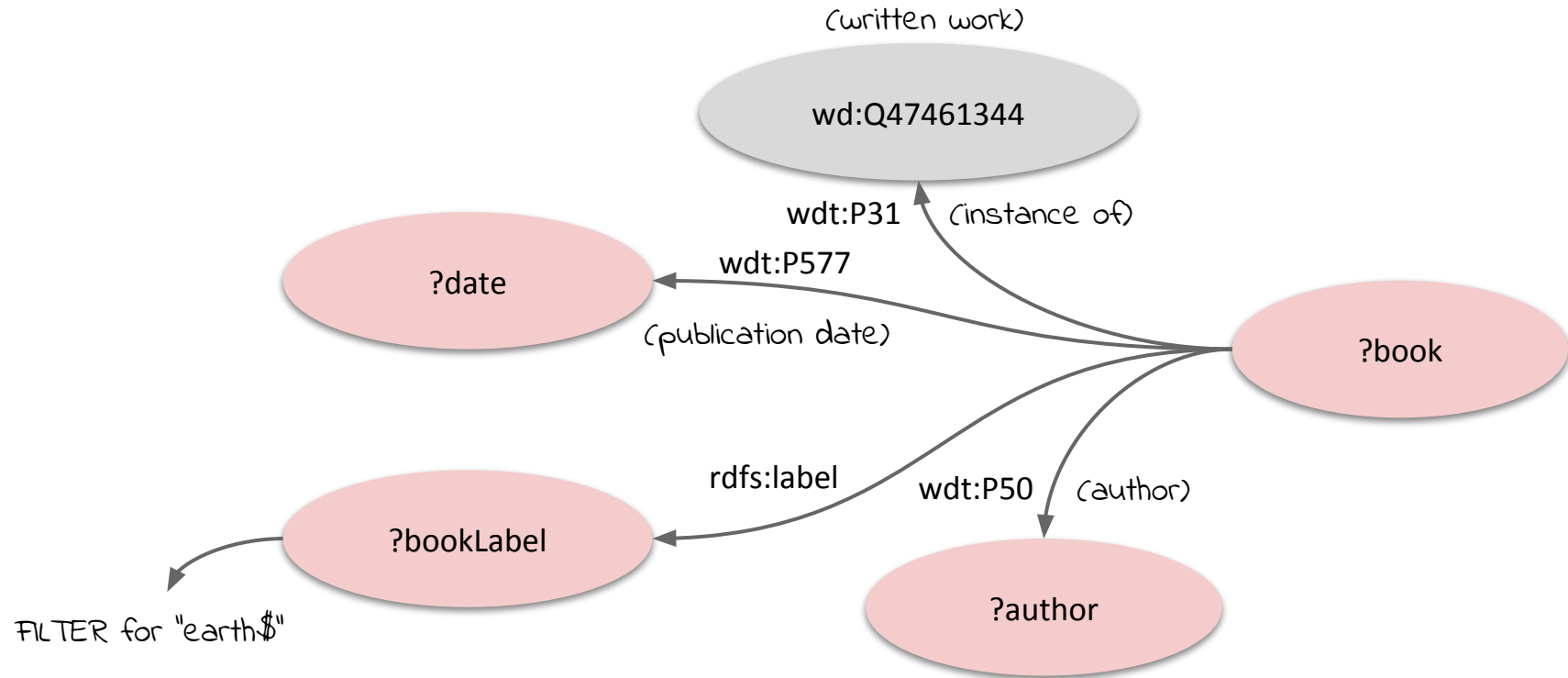
More SPARQL Operators

- Logical connectives `&&` (**AND**) and `||` (**OR**) for xsd:boolean
- Comparison operators `=`, `!=`, `<`, `>`, `<=`, and `>=` for numeric datatypes, xsd:dateTime, xsd:string, and xsd:boolean
- Comparison operators `=` and `!=` for other datatypes
- Arithmetic operators `+`, `−`, `*`, and `/` for numeric datatypes

- and in addition:
    - `REGEX(String,Pattern)` or `REGEX(String,Pattern,Flags)`
    - `sameTERM(A,B)`
    - `langMATCHES(A,B)`

SPARQL Filter Constraints

- What book titles end with the word "**earth**" sorted by publication date?

SPARQL Filter Constraints

- What book titles end with the word "**earth**" sorted by publication date?

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?authorLabel ?bookLabel ?date
WHERE {
    ?book wdt:P31 wd:Q47461344 .
    ?book wdt:P50 ?author .
    ?book wdt:P577 ?date .
    ?book rdfs:label ?bookLabel
    FILTER (LANG(?bookLabel)="en")
    FILTER REGEX (?bookLabel,"earth$","i") .
    ?author rdfs:label ?authorLabel
    FILTER (LANG(?authorLabel)="en") .
} ORDER BY ?date
```

string

regular expression

flags

- With **FILTER REGEX**, regular expressions can be filtered

query SPARQL endpoint

10

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

- What book titles end with the word "earth" sorted by publication date?

Wikidata Query Service

```
1  PREFIX wd: <http://www.wikidata.org/entity/>
2  PREFIX wdt: <http://www.wikidata.org/prop/direct/>
3  PREFIX wikibase: <http://wikiba.se/ontology#>
4  PREFIX bd: <http://www.bigdata.com/rdf#>
5
6  SELECT ?authorLabel ?bookLabel ?date
7  WHERE {
8      ?book wdt:P31 wd:Q47461344 .
9      ?book wdt:P50 ?author .
10     ?book wdt:P577 ?date .
11     ?book rdfs:label ?bookLabel FILTER (LANG(?bookLabel)="en")
12     FILTER regex (?bookLabel,"earth$","i") .
13     ?author rdfs:label ?authorLabel FILTER (LANG(?authorLabel)="en")
14 }
15 ORDER BY ?date
16
```
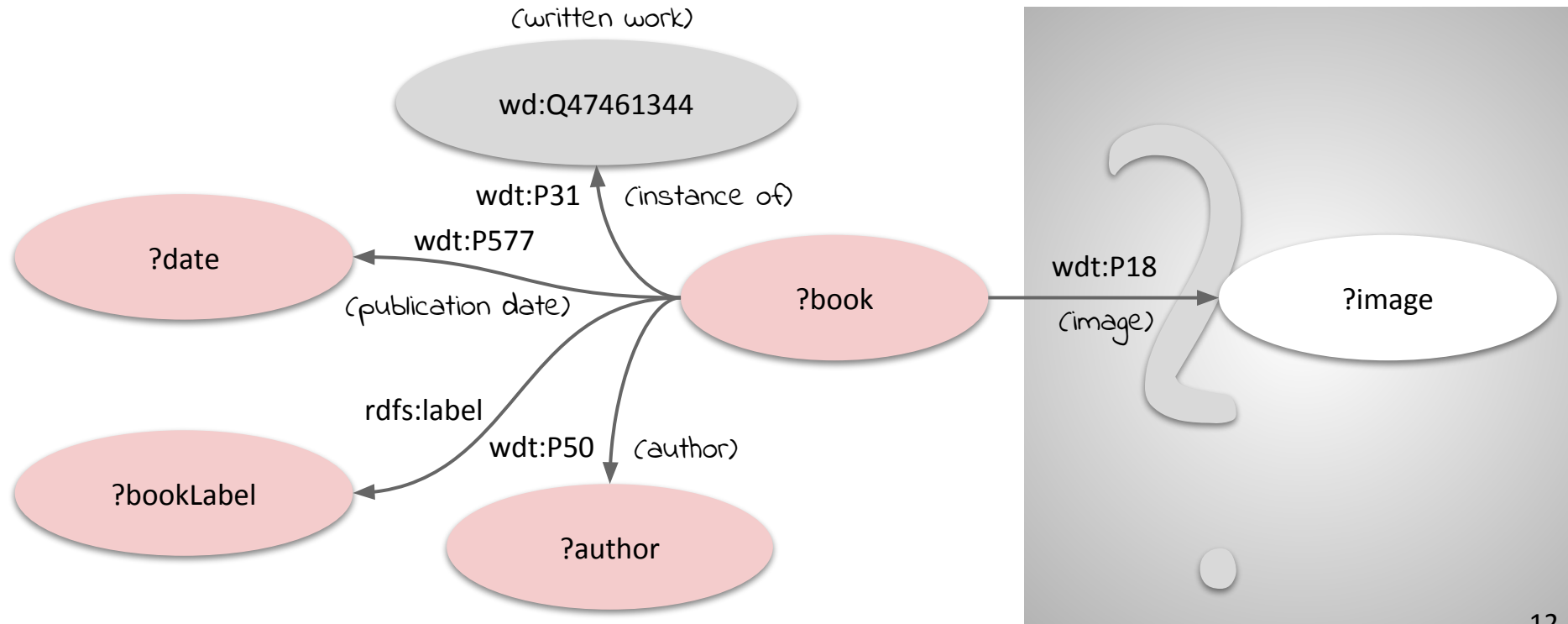
46 results in 5623 ms

authorLabel	bookLabel	date
Charles Dickens	The Cricket on the Hearth	20 December 1845
André Gide	The Fruits of the Earth	1 January 1897
H. G. Wells	The Food of the Gods and How It Came to Earth	1 January 1904
Bruce Marshall	Children of This Earth	1 January 1930
Pearl S. Buck	The Good Earth	2 March 1931
Raymond F. Jones	This Island Earth	1 January 1952
François Bordes	Fleeing Earth	1 January 1960
Arthur Koestler	Scum of the Earth	1 January 1968
Larry Niven	A Gift from Earth	1 January 1968
Robert Silverberg	Downward to the Earth	1 January 1970
Hal Lindsey	The Late, Great Planet Earth	1 January 1970
Robert Foster	The Complete Guide to Middle-earth	1 January 1978

query SPARQL endpoint

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

SPARQL Filter Constraints

- Which book titles **end with the word "earth", and, if available, do also have an image?**

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

SPARQL Filter Constraints

- Which book titles **end with the word "earth"**, and, if available, **do also have an image?**

- Optional selection of graph pattern via **OPTIONAL**

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?authorLabel ?bookLabel ?date ?image
WHERE {
    ?book wdt:P31 wd:Q47461344 .
    ?book wdt:P50 ?author .
    ?book wdt:P577 ?date .
    ?book rdfs:label ?bookLabel
    FILTER (LANG(?bookLabel)="en")
    FILTER regex (?bookLabel,"earth$","i") .
    ?author rdfs:label ?authorLabel
    FILTER (LANG(?authorLabel)="en") .
    OPTIONAL {?book wdt:P18 ?image .}
} ORDER BY ?date
```

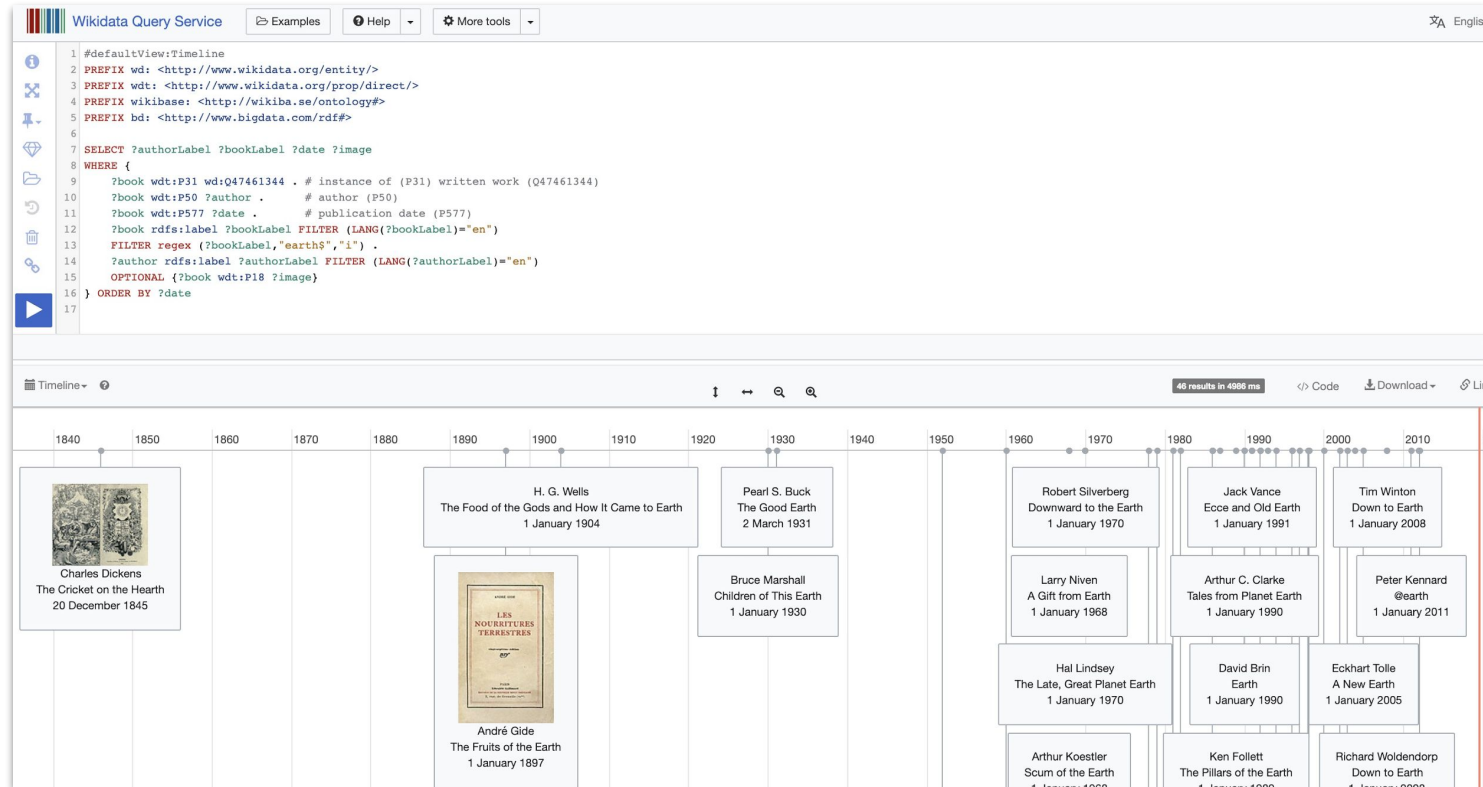*optional constraint*
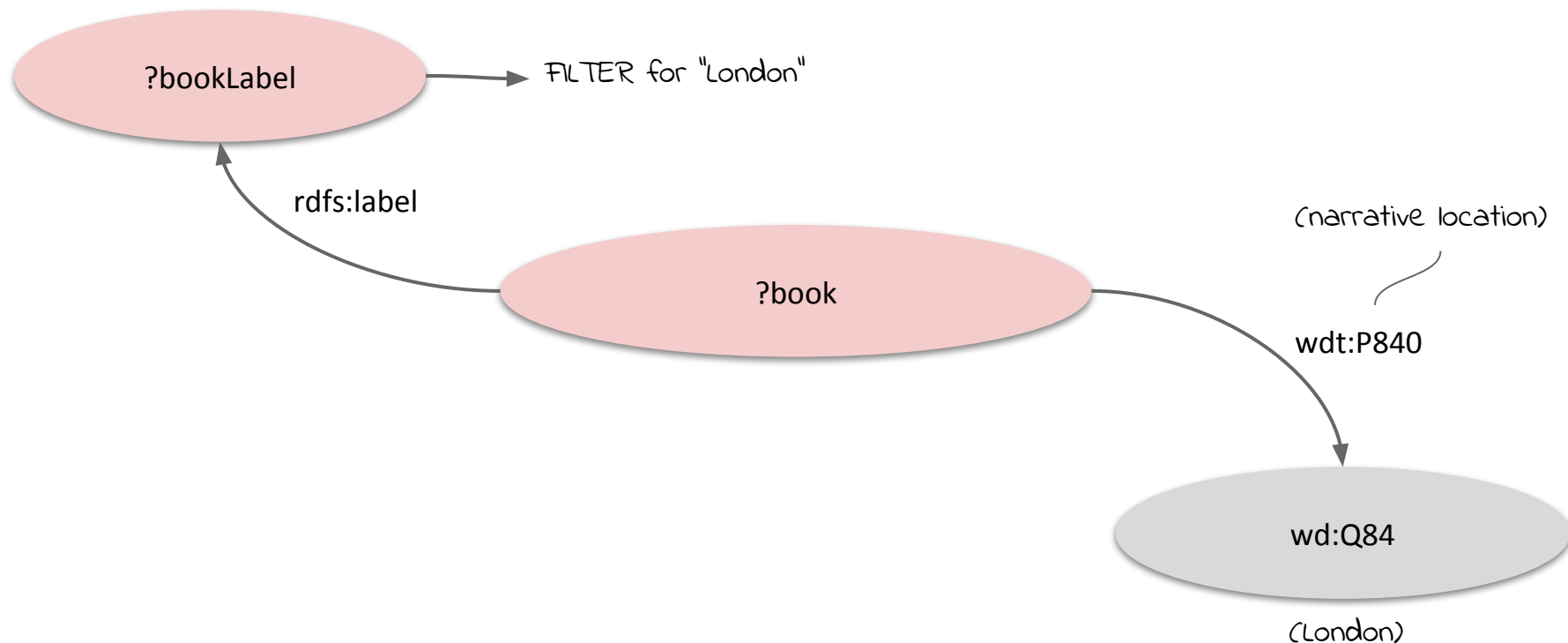
[query SPARQL endpoint](#)

- Which book titles **end with the word "earth", and, if available, do also have an image?**

[query SPARQL endpoint](#)

SPARQL Alternative Results via UNION

- Example: which books **mention "London" in their title** or

  have **London as their narrative location**

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

SPARQL Alternative Results via UNION

- Example: which books **mention "London" in their title or**

   have **London as their narrative location**

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX bd: <http://www.bigdata.com/rdf#>


SELECT ?authorLabel ?bookLabel ?book ?date ?image
WHERE {
    ?book wdt:P31 wd:Q47461344 .
    ?book wdt:P50 ?author .
    ?book wdt:P577 ?date .
    { FILTER regex (?bookLabel,"London","i") . }
    UNION
    { ?book wdt:P840 wd:Q84 .}
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
} ORDER BY ?date
```

logical disjunction

- The keyword **UNION** allows for alternatives (logical disjunction)

query SPARQL endpoint

- Example: which books mention "London" in their title or

   have London as their narrative location


```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX bd: <http://www.bigdata.com/rdf#>

SELECT ?authorLabel ?bookLabel ?book ?date
WHERE {
    ?book wdt:P31 wd:Q47461344 . # instance of (P31) written work (Q47461344)
    ?book wdt:P50 ?author .      # author (P50)
    ?book wdt:P577 ?date .       # publication date (P577)
    {
     FILTER regex (?bookLabel,"London","i") .
    }
    UNION
    {
     ?book wdt:P840 wd:Q84       # narrative location (P840) London (Q84)
    }
    SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
} ORDER BY ?date
```
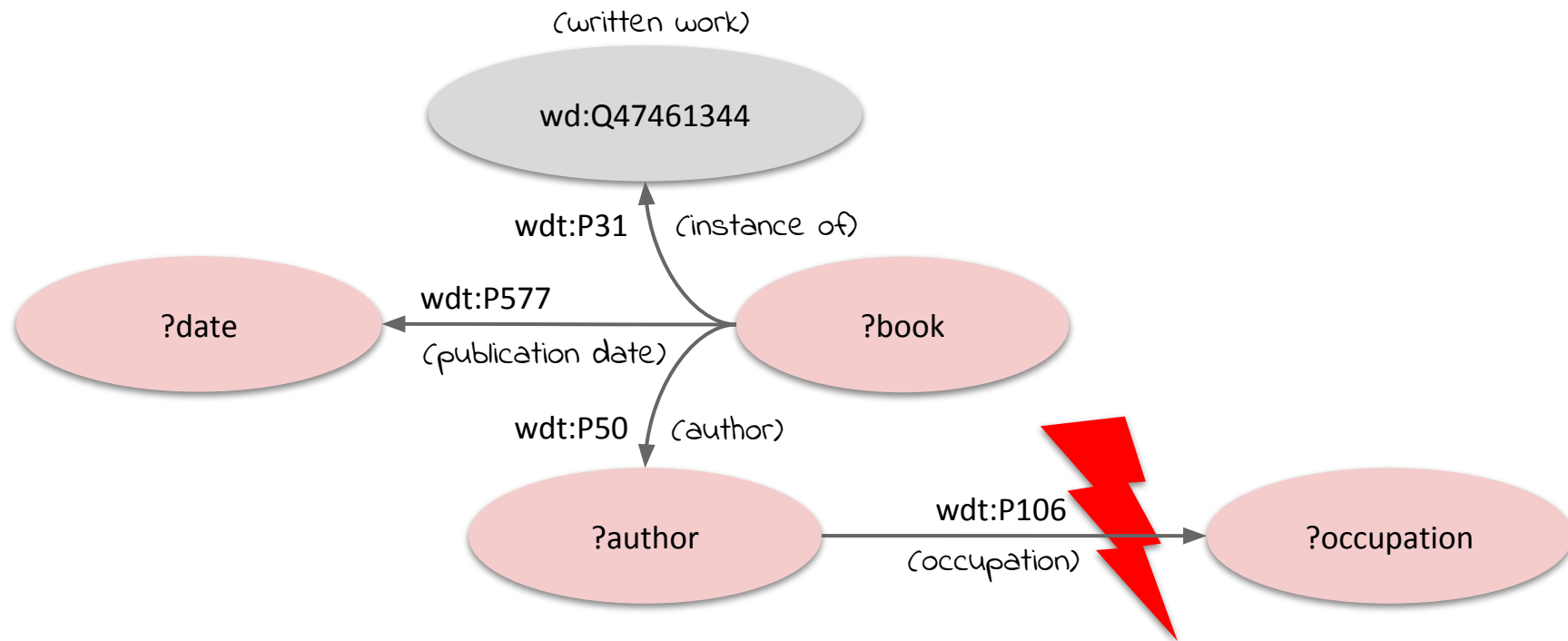
238 results in 1721 ms

authorLabel	bookLabel	book	date
Girolamo Graziani	Il Cromuele	wd:Q3792648	1 January 1671
Daniel Defoe	A Journal of the Plague Year	wd:Q1215399	1 January 1722
Frances Burney	Cecilia	wd:Q3233990	1 July 1782
Charles Dickens	Nicholas Nickleby	wd:Q847642	1 January 1839
Charles Dickens	A Christmas Carol	wd:Q62879	19 December 1843
Alexandre Dumas	The Three Musketeers	wd:Q140527	1 January 1844
Charles Dickens	The Cricket on the Hearth	wd:Q825220	20 December 1845
Charles Dickens	A Tale of Two Cities	wd:Q308918	1 January 1859
Charles Dickens	Great Expectations	wd:Q219552	1 January 1860

query SPARQL endpoint

17

Knowledge Graphs 2020 , Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

SPARQL Negation

- Example: which books are written by authors who **don't have an occupation**?

(written work)

wd:Q47461344

wdt:P31 (instance of)

wdt:P577 (publication date)

?date

?book

wdt:P50 (author)

?author

wdt:P106 (occupation)

?occupation

SPARQL Negation

- Example: which books are written by authors who **don't have an occupation**?

```
SELECT ?authorLabel ?bookLabel ?date
WHERE {
    ?book wdt:P31 wd:Q47461344 .
    ?book wdt:P50 ?author .
    FILTER NOT EXISTS {?author wdt:P106 ?occupation }
?book wdt:P577 ?date .
    SERVICE wikibase:label
    { bd:serviceParam wikibase:language "en, de, es, it" }
}
```

filter query result for existence

▸ SPARQL 1.1 offers several variants for negation:

- ○ **FILTER NOT EXISTS**
- ○ **MINUS**
- ○ **!BOUND()**

query SPARQL endpoint

- Example: which books are written by authors who **don't have an occupation** ?


```
1  SELECT ?authorLabel ?bookLabel ?date
2  WHERE {
3      ?book wdt:P31 wd:Q47461344 .  # instance of (P31) written work (Q47461344)
4      ?book wdt:P50 ?author .        # author (P50)
5      FILTER NOT EXISTS {?author wdt:P106 ?occupation }
6      ?book wdt:P577 ?date .          # publication date (P577)
7      SERVICE wikibase:label
8      { bd:serviceParam wikibase:language "en, de, es, it" }
9  }
10
```

509 results in 1928 ms

authorLabel	bookLabel	date
various authors	Allgemeine Deutsche Biographie	1 January 1875
various authors	The Merck Index	1 January 1889
World Organization of Family Doctors	International Classification of Primary Care	1 January 1987
Institute of the Italian Encyclopaedia	Enciclopedia Treccani	1 January 1929
Institute of the Italian Encyclopaedia	Enciclopedia Treccani	1 January 1939
Fruttero & Lucentini	The Sunday Woman	1 January 1972
Monaldi & Sorti	Imprimatur	1 January 2002
various authors	The Banksia Atlas	1 January 1988
Sjöwall and Wahlöö	The Laughing Policeman	1 January 1968
Sjöwall and Wahlöö	The Abominable Man	1 January 1971

query SPARQL endpoint

More Complex SPARQL Queries

21

[4]

Picture References:

- [1] Benjamin Nowack, *The Semantic Web - Not a Piece of cake…*, at bnode.org, 2009-07-08 , [CC BY 3.0]
  http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake

- [2] British Crown vector illustration, publicdomainvectors.org, [Public Domain]
  https://publicdomainvectors.org/en/free-clipart/British-Crown-vector-illustration/12150.html

- [3] WIkidata Logo, Planemad, [Public Domain]
  https://commons.wikimedia.org/wiki/File:Wikidata-logo-en.svg

- [4] Ernst Haeckel, Kunstformen der Natur (1904), plate 54: Gamochonia, [Public Domain]
  https://commons.wikimedia.org/wiki/File:Haeckel_Gamochonia.jpg