SPARQL, a query language for RDF databases

Silvio Peroni

silvio.peroni@unibo.it - https://orcid.org/0000-0003-0530-4305 - @essepuntato

<u>Data Science (A.Y. 2023/2024)</u> <u>Second Cycle Degree in Digital Humanities and Digital Knowledge</u> <u>Alma Mater Studiorum - Università di Bologna</u>





Summary of the previous lectures (1/2)

A datum is a declarative statement subject-predicate-object that, through the predicate, either attributes a literal (i.e. a value such as a string, a number, etc.) to a subject entity or it relates such a subject entity with another entity

Each entity, being used either as subject or object of a statement, is characterised by a unique identifier

The **same entity** can be used as **subject** or object in one or more data, while a literal **cannot be used** as **subject** in any datum

An attribute is intrinsically **part of** the **entity** to which it is associated – modifying the value of an attribute affect **only** the **entity** to which it refers to

A **data model** is an abstract, simplified and formal representation of some data related to a system or a real domain, and enables us to describe what a data collection is about and to check data correctness

A data model permit one to specify classes of entities, their attributes and relations

Summary of the previous lectures (2/2)

Depending on the structure in which data are stored (or exposed), you need to approach the queries to datasets from a different angle

- With **tabular data**, often you have to combine tables between them to obtain bigger tables which contain the query requirements and the related answer
- With **graph data**, you explore the graph starting from fixed points (i.e. known entities, values, predicates) to find a pattern that is compliant with the query

A database as a collection of data which organised, stored and accessed electronically, which can be created through a database management system (DBMS)

A **transaction** is a unit of work performed (compliant with **ACID properties**) within a DBMS against a database and usually represents any change in a database

SQL is a **query language** used and designed for managing data in a **relational** database management system, and allows one to **create** tables, to **populate** them, and finally to **query** them using a particular syntax

Any question about the previous lecture?

SPARQL

The SPARQL Protocol and RDF Query Language (SPARQL, and yes: it is a recursive name) is a query language used and designed for managing RDF data, as those available in an RDF triplestore

It is a standard that is implemented in all RDF-based triplestore, and allows one to populate and query them using a particular syntax

The full specification of the language is a W3C Recommendation and it is entirely available at https://www.w3.org/TR/sparql11-query/

A good tutorial is available at:

https://docs.data.world/tutorials/sparql/primer_intro.html

Basic SPARQL syntax for queries

```
SELECT <variables>
WHERE {
    <triple pattern 1> .
    <triple pattern 2> .
...
}
```

A space-separated list of ?x variables to have as result of the query ("*" means all the variables)

A series of conditions involving the ?x variables

... triple patterns **s p o** where:

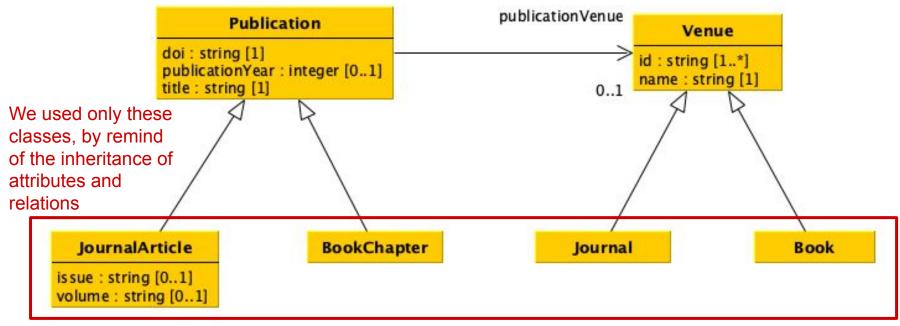
above used in one or more...

- s can be a <url>
 <url>
 <url> resource or a ?x
- P can be a <url> property or a ?x variable
- o can be a <url> resource, a
 "literal" or a ?x variable

Questions to remind for designing a query:

- Which resources are known?
- Which literals are known?
- Which properties are known?

Exemplar classes and properties for queries



Remind that we used strings to store years of publication instead of integers

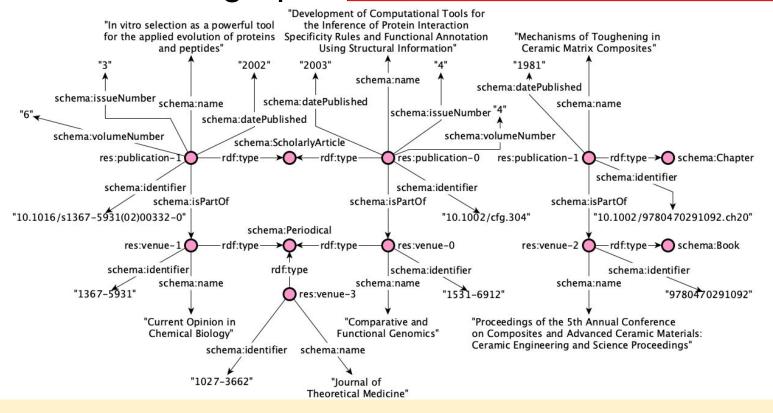
Data in the RDF graph

Prefixes

PREFIX res: https://comp-data.github.io/res/

PREFIX rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

PREFIX schema: https://schema.org/>



Queries

Retrieve complete information about all journal articles

Retrieve the titles of all journal articles

Retrieve the title of the journal article with DOI "10.1016/s1367-5931(02)00332-0"

Retrieve the title of the publication with DOI "10.1016/s1367-5931(02)00332-0"

Return the name of the journal of the article with DOI "10.1016/s1367-5931(02)00332-0"

Return the id and name of the journal of the article with DOI "10.1016/s1367-5931(02)00332-0"

Retrieve complete information about all journal articles

- Which resources are known?
- Which literals are known?
- Which properties are known?

```
WHERE
  ?s rdf:type
  ?s ?p ?o .
```

```
"Development of Computational Tools for
                                                                                the Inference of Protein Interaction
                                           "In vitro selection as a powerful tool
                                           for the applied evolution of proteins
                                                                            Specificity Rules and Functional Annotation
                                                                                                                        "Mechanisms of Toughening in
                                                     and peptides"
                                                                                ?o Using Structural Information"
                                                                                                                         Ceramic Matrix Composites"
                                                                    "2002"
                                                                                "2003"
                                                                                                                        "1981"
                                                                                          schema:name
                                                                     ?p
                                                                                                                 schema:datePublished
                                 schema:issueNumber schema:name
                          ?o
                                                                        schema: datePublished
                                                                                                                                schema:name
                                                                                                schema:issueNumber "4" ?0
                                                ?p
                                                             schema:datePublished
                                                                                                        ?p chema:volumeNumber
                                 schema:volumeNumber
                                                                   schema:ScholarlyArticle
                                          res:publication-1 ?s
                                                                           ?o ← rdf:type
                                                                                               ?s res:publication-0
                                                                 rdf:type-
                                                                                                                     res:publication-1
                                                                                                                                         rdf:type > chema:Chapter
                                                                                                                                       schema:identifier
                                           schema:identifier
                                                                                                 schema:identifier
                             ?0
                                                    schema:isPartOf
                                                                                         schema:isPartOf
                                                                                                                    ?0
                                                                                                                               schema:isPartOf
                          "10.1016/s1367-5931(02)00332-0 2n
                                                                                                          "10.1002/cfg.304"
                                                                                                                                       "10.1002/9780470291092.ch20"
                                                                     schema:Periodical
                                                                                               ?o res:venue-0
                                              res:venue-1?0
                                                                                                                         res:venue-2 rdf:type > schema:Book
                                                                           → rdf:type-
                                           schema:identifier
                                                                          rdf:type
                                                                                                schema:identifier
                                                                                                                                          schema:identifier
                                                     schema:name
                                                                                          schema:name
                                                                                                                                schema:name
                                     "1367-5931"
                                                                             res:venue-3
                                                                                                          "1531-6912"
                                                                                                                                                     "9780470291092"
                                                                                                                "Proceedings of the 5th Annual Conference
                                                  "Current Opinion in
                                                                                        "Comparative and
                                                   Chemical Biology'
                                                                                      Functional Genomics"
                                                                                                                on Composites and Advanced Ceramic Materials:
                                                                                                                Ceramic Engineering and Science Proceedings"
                                                              schema:identifier
                                                                               schema:name
                                                           "1027-3662"
                                                                                      "Journal of
                                                                                 Theoretical Medicine'
schema:ScholarlyArticle .
```

Retrieve the titles of all journal articles

- Which resources are known?
- Which literals are known?
- Which properties are known?

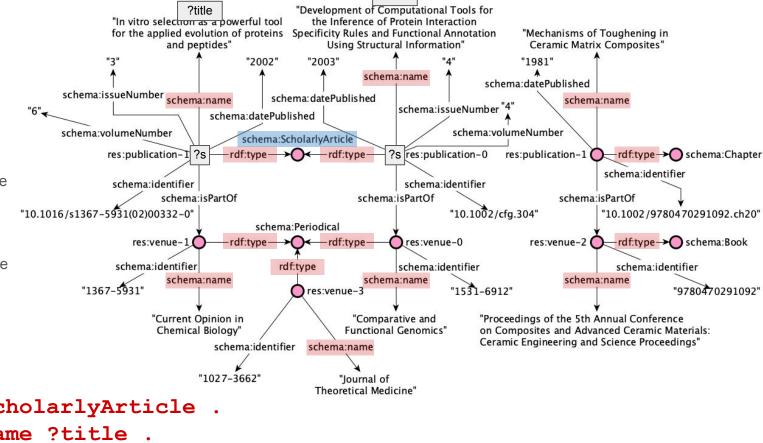
```
SELECT ?title
```

WHERE

?s rdf:type

schema: ScholarlyArticle .

?s schema:name ?title .



?title

Retrieve the title of the journal article with DOI "10.1016/s1367-5931(02)00332

- Which resources are known?
- Which literals are known?
- Which properties are known?

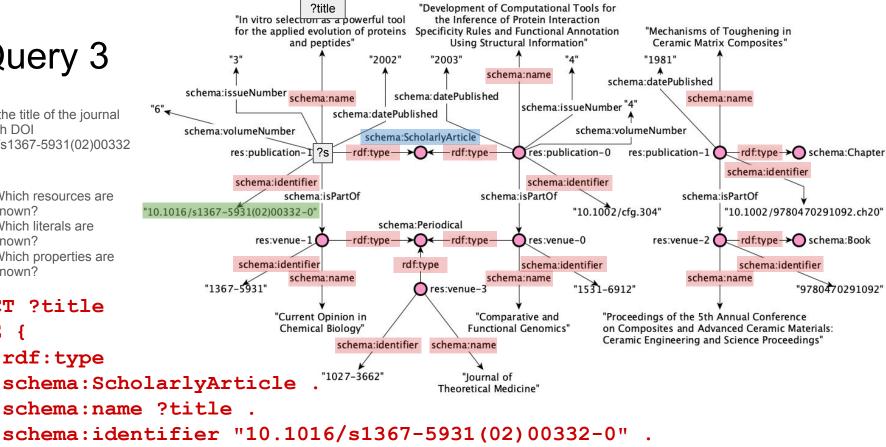
```
SELECT ?title
```

```
WHERE
```

?s rdf:type

?s schema:name ?title

?s schema:identifier "10.1016/s1367-5931(02)00332-0"



Retrieve the title of the publication with DOI "10.1016/s1367-5931(02)0 0332-0"

- Which resources are known?
- Which literals are known?
- Which properties are known?

```
"Current Opinion in
                                         Chemical Biology'
SELECT ?title
                                                schema:identifier
                                                           schema:name
WHERE
  VALUES ?type {
                                              "1027-3662"
                                                               "lournal of
     schema:ScholarlyArticle schema:Chapter }
  ?s rdf:type ?type .
      schema: name ?title .
  ?s schema:identifier "10.1016/s1367-5931(02)00332-0" . }
```

```
?title
                                                      "Development of Computational Tools for
                  "In vitro selection as a powerful tool
                                                         the Inference of Protein Interaction
                 for the applied evolution of proteins Specificity Rules and Functional Annotation
                                                                                                  "Mechanisms of Toughening in
                            and peptides"
                                                            Using Structural Information"
                                                                                                   Ceramic Matrix Composites"
                                            "2002"
                                                                                                  "1981"
                                                        "2003"
                                                                   schema:name
                                                                                           schema:datePublished
       schema: is sue Number \\ \underline{schema: name}
                                                schema: datePublished
                                                                                                          schema:name
                                                                         schema:issueNumber "4
                                    schema:datePublished
        schema:volumeNumber
                                                                                    schema:volumeNumber
                                           schema:ScholarlyArticle
                res:publication-1 ?s
                                        -rdf:type → >
                                                          -rdf:type-
                                                                         res:publication-0
                                                                                               res:publication-1
                                                                                                                    -rdf:type → schema:Chapter
                                                   ?type
                                                                                                                                        ?type
                                                                                                                  schema:identifier
                 schema:identifier
                                                                           schema:identifier
                           schema:isPartOf
                                                                                                         schema:isPartOf
                                                                  schema:isPartOf
"10.1016/s1367-5931(02)00332-0"
                                                                                    "10.1002/cfg.304"
                                                                                                                  "10.1002/9780470291092.ch20"
                                             schema:Periodical
                                                                                                                   rdf:type > schema:Book
                      res:venue-1
                                                   → rdf:type-
                                                                         res:venue-0
                                                                                                   res:venue-2
                  schema:identifier
                                                                         schema:identifier
                                                  rdf:type
                                                                                                                    schema:identifier
                                                                                                          schema:name
                            schema:name
                                                                   schema:name
            "1367-5931"
                                                                                    "1531-6912"
                                                                                                                                "9780470291092"
                                                     nes:venue-3
                                                                                          "Proceedings of the 5th Annual Conference
                                                                 "Comparative and
                                                               Functional Genomics"
                                                                                          on Composites and Advanced Ceramic Materials:
                                                                                          Ceramic Engineering and Science Proceedings"
                                                         Theoretical Medicine"
```

Return the name of the journal of the article with DOI "10.1016/s1367-5931(02)00 332-0"

- Which resources are known?
- Which literals are known?

?name

SELECT

WHERE

Which properties are known?

```
"Development of Computational Tools for
                                                 "In vitro selection as a powerful tool
                                                                                    the Inference of Protein Interaction
                                                 for the applied evolution of proteins
                                                                                 Specificity Rules and Functional Annotation
                                                                                                                          "Mechanisms of Toughening in
                                                          and peptides"
                                                                                       Using Structural Information"
                                                                                                                           Ceramic Matrix Composites"
                                                                         "2002"
                                                                                                                          "1981"
                                                                                    "2003"
                                                                                              schema:name
                                                                                                                   schema:datePublished
                                        schema: is sue Number \\ \underline{schema: name}
                                                                             schema: datePublished
                                                                                                                                 schema:name
                                                                                                    schema:issueNumber "4
                                                                  schema:datePublished
                                        schema:volumeNumber
                                                                                                              schema:volumeNumber
                                                                        schema:ScholarlyArticle
                                                res:publication-1 ?s
                                                                     -rdf:type---rdf:type-
                                                                                                   res:publication-0
                                                                                                                       res:publication-1
                                                                                                                                          -rdf:type → schema:Chapter
                                                                                                                                         schema:identifier
                                                 schema:identifier
                                                                                                     schema:identifier
                                                          schema:isPartOf
                                                                                             schema:isPartOf
                                                                                                                                 schema:isPartOf
                                "10.1016/s1367-5931(02)00332-0"
                                                                                                             "10.1002/cfg.304"
                                                                                                                                        "10.1002/9780470291092.ch20"
                                                                          schema:Periodical
                                                    res:venue?journal_rdf:tyne
                                                                                                                                          rdf:type > chema:Book
                                                                                  rdf:type
                                                                                                     res:venue-0
                                                                                                                           res:venue-2
                                                 schema:identifier
                                                                                                    schema:identifier
                                                                              rdf:type
                                                                                                                                           schema:identifier
                                                                                                                                  schema:name
                                                          schema:name
                                                                                              schema:name
                                           "1367-5931"
                                                                                                             "1531-6912"
                                                                                                                                                     "9780470291092"
                                                                                 nes:venue-3
                                                                                                                   "Proceedings of the 5th Annual Conference
                                                        "Current Opinion in
                                                                                            "Comparative and
                                                         Chemical Biology"
                                                                                          Functional Genomics"
                                                                                                                   on Composites and Advanced Ceramic Materials:
                                                                                                                   Ceramic Engineering and Science Proceedings"
                                                           ?name schema:identifier
                                                                                    schema:name
                                                                "1027-3662"
                                                                                          "Journal of
                                                                                     Theoretical Medicine"
   rdf:type schema:ScholarlyArticle .
     schema:identifier "10.1016/s1367-5931(02)00332-0".
?s schema:isPartOf ?journal .
?journal schema:name ?name . }
```

Return the id and name of the journal of the article with DOI

"10.1016/s1367-5931(02)00 332-0"

- Which resources are known?
- Which literals are known?

SELECT

WHERE {

Which properties are known?

?name ?id

```
"In vitro selection as a powerful tool
                                                                                    the Inference of Protein Interaction
                                                for the applied evolution of proteins
                                                                                Specificity Rules and Functional Annotation
                                                                                                                         "Mechanisms of Toughening in
                                                          and peptides"
                                                                                      Using Structural Information"
                                                                                                                          Ceramic Matrix Composites"
                                                                        "2002"
                                                                                                                         "1981"
                                                                                   "2003"
                                                                                             schema:name
                                                                                                                  schema:datePublished
                                        schema: is sue Number\\ schema: name
                                                                            schema: datePublished
                                                                                                                                schema:name
                                                                                                   schema:issueNumber "4
                                                                  schema:datePublished
                                        schema:volumeNumber
                                                                                                             schema:volumeNumber
                                                                        schema:ScholarlyArticle
                                               res:publication-1 ?s
                                                                     -rdf:type---rdf:type-
                                                                                                  res:publication-0
                                                                                                                      res:publication-1
                                                                                                                                         -rdf:type -> schema:Chapter
                                                                                                                                       schema:identifier
                                                schema:identifier
                                                                                                    schema:identifier
                                                         schema:isPartOf
                                                                                            schema:isPartOf
                                                                                                                               schema:isPartOf
                                "10.1016/s1367-5931(02)00332-0"
                                                                                                            "10.1002/cfg.304"
                                                                                                                                       "10.1002/9780470291092.ch20"
                                                                          schema:Periodical
                                                    res:venue?journal_rdf:tyne
                                                                                                                                         rdf:type > chema:Book
                                                                                 rdf:type
                                                                                                   res:venue-0
                                                                                                                          res:venue-2
                                                 schema:identifier
                                                                                                   schema:identifier
                                                                              rdf:type
                                                                                                                                         schema:identifier
                                                                                                                                schema:name
                                                          schema:name
                                                                                             schema:name
                                           "1367-5931"
                                                                                                            "1531-6912"
                                                                                                                                                    "9780470291092"
                                                                                nes:venue-3
                                                                                                                  "Proceedings of the 5th Annual Conference
                                                        "Current Opinion in
                                                                                           "Comparative and
                                                        Chemical Biology"
                                                                                         Functional Genomics"
                                                                                                                 on Composites and Advanced Ceramic Materials:
                                                                                                                 Ceramic Engineering and Science Proceedings"
                                                          ?name schema:identifier
                                                                                   schema:name
                                                                "1027-3662"
                                                                                         "lournal of
?s rdf:type schema:ScholarlyArticle .
                                                                                    Theoretical Medicine"
    schema:identifier "10.1016/s1367-5931(02)00332-0".
    schema:isPartOf ?journal .
?journal schema:name ?name
?journal schema:identifier ?id . }
```

"Development of Computational Tools for

Do you want to try them with real data?

You should have already downloaded the .jar file of Blazegraph for the previous lecture

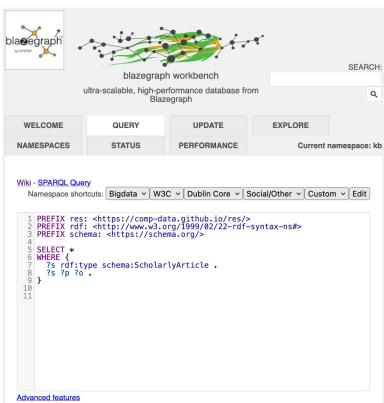
(https://github.com/comp-data/2021-2022/tree/main/docs/handson/05)

Start Blazegraph from the terminal with java -server -Xmx1g -jar blazegraph.jar

Open your browser and go to http://127.0.0.1:9999/blazegraph/

Select the tab "Query" and write and execute SPARQL queries (remember the prefixes if you want to use them in the query)

Close Blazegraph from the terminal pressing CTRL+C on your keyboard



End

SPARQL, a query language for RDF databases

Silvio Peroni

<u>silvio.peroni@unibo.it</u> – <u>https://orcid.org/0000-0003-0530-4305</u> – <u>@essepuntato</u>

<u>Data Science (A.Y. 2023/2024)</u> <u>Second Cycle Degree in Digital Humanities and Digital Knowledge</u> <u>Alma Mater Studiorum - Università di Bologna</u>



