Group A3D: SPARQL queries and Analytics

Group members: Andrea Bruttomesso 2120933
Alessandro Corrò 2125034
Davide Seghetto 2122548
Andrea Stocco 2108885

TASK

Provide at least 8 SPARQL queries over your RDF datasets. You may also perform advanced data analytics to uncover interesting insights from your datasets. Please submit a PDF that includes the SPARQL queries along with relevant plots or tables summarizing your analytics. For each query, provide a description that explains its purpose and overall objective.

Query 1 - papersNobelTopicsYear

This query shows the topics present in both Nobel motivations and paper abstracts. For a given year, it returns the number of paper in which these topics appear. This query can be used to find correlations between Nobel Prize topics and research papers.

For instance, running this query for the year 2004 shows that the topic "protein" appeared in 28 papers. Hence, we could say that in 2004 chemistry was the main research field (considering the limited number of papers available) and that the main topic of the Nobel Prize awarded in that year was studied by several researcher.

Unfortunately, this query is not always useful. In some cases, the main topics may include words like "method" and "analysis", which are not informative enough to determine how extensively a specific topic was studied in a given year.

Due to the distribution of research papers in our dataset across different years, this query provides more meaningful results for years after 2000.

```sparql
PREFIX spif: <http://spinrdf.org/spif#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX jur: <http://sweet.jpl.nasa.gov/2.3/humanJurisdiction.owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?singleAbTopic ?nobel (COUNT(?paper) AS ?numPapers) WHERE {
    {
        SELECT ?singleAbTopic ?paper WHERE {
            ?paper :hasAbstractTopics ?topics;
                   :hasYear "2014"^^xsd:gYear.
            ?singleAbTopic spif:split(?topics ",")
        }
    }
    {
        SELECT ?singleNoTopic ?nobel WHERE {
            ?nobel :hasMotivationTopics ?topics;
                   :hasYear "2014"^^xsd:gYear.
            ?singleNoTopic spif:split(?topics ",")
        }
    }
    FILTER (?singleAbTopic = ?singleNoTopic)
}
GROUP BY ?singleAbTopic ?nobel
ORDER BY desc(?numPapers)
```

Query 2 - papersPerTopic

```
1  PREFIX spif: <http://spinrdf.org/spif#>
2  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3  PREFIX jur: <http://sweet.jpl.nasa.gov/2.3/humanJurisdiction.owl#>
4  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
6  PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
7  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
8
9  SELECT ?singleTopic (COUNT(?paper) AS ?numPapers) WHERE {
10     ?paper :hasAbstractTopics ?topics.
11     ?singleTopic spif:split(?topics ",")
12 }
13 GROUP BY ?singleTopic
14 ORDER BY desc(?numPapers)
```

Papers per topic

Query 3 - sharedNobels

This query shows the number of Nobel Prizes shared by multiple laureates and the number of laureates sharing Nobel Prizes.

The query provides an interesting result: 242 out of 579 Nobel Prizes (41.8%) have been shared by multiple laureates, and 632 laureates have shared different Nobel Prizes. On average, a Nobel Prize is shared by more than 2 laureates (2.6 laureates per prize).

```
1  PREFIX spif: <http://spinrdf.org/spif#>
2  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3  PREFIX jur: <http://sweet.jpl.nasa.gov/2.3/humanJurisdiction.owl#>
4  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
6  PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
7  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
8
9  SELECT (COUNT(?nobel) AS ?numSharedNobels) (SUM(?share) AS numLaureatesSharingNobels)
       WHERE {
10         ?nobel :hasPrizeShare ?share.
11         FILTER (?share > 1)
12 }
```

how many nobels are shared? how many laureates are sharing a nobel?

Query 4 - laureatesCollaborations

```
1  PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
2  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
4
5  SELECT ?title (GROUP_CONCAT(?name; separator=", ") AS ?laureates) WHERE {
6    ?laureate rdf:type :Laureate .
7    ?paper rdf:type :Paper ;
8          :hasTitle ?title .
9    ?laureate :hasWritten ?paper .
10   ?laureate foaf:name ?name .
11 }
12 GROUP BY ?title
13 HAVING (COUNT(DISTINCT ?laureate) > 1)
```

Laureates that wrote a paper together

Query 5 - laureatesPerCountryAndFunding

```
1  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
2  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3  PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
4
5  SELECT ?year ?state (COUNT(?laureate) AS ?howmany) ?amount WHERE {
6      ?laureate :hasWon ?nobel .
7      ?nobel :hasYear ?year .
8      ?laureate :basedIn | :bornIn ?city .
9      ?city :locatedIn ?country .
10     ?country foaf:name ?state .
11     OPTIONAL {
12         ?country :hasFunded ?funding .
13         ?funding a :Funding ;
14           :hasYear ?year ;
15                 :hasAmount ?amount .
16     }
17 }
18 GROUP BY ?year ?state ?amount
19 ORDER BY DESC (?howmany)
```

For every year and country, the query extracts the number of laureate of that country and, if available, the funding for R&D : how fundings in R&D affect the possibility to win?

Query 6 - moreThanOneNobel

```
1  PREFIX spif: <http://spinrdf.org/spif#>
2  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3  PREFIX jur: <http://sweet.jpl.nasa.gov/2.3/humanJurisdiction.owl#>
4  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
6  PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
7  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
8
9  SELECT ?laureate (COUNT(?nobel) AS ?numNobels) WHERE {
10     ?laureate :hasWon ?nobel.
11 }
12 GROUP BY ?laureate
13 HAVING (?numNobels > 1)
```

who won more than one nobel prize

Query 7 - papersPerVenue

This plot shows the number of papers published over the years by major venues (those with at least 800 papers published, according to our dataset).

In recent years, Bioinformatics could be considered one of the most influential venue due to its consistently higher number of papers published compared to others. IEEE venues, are the most prominent in the fields of information and tecnology.

For instance, on 2009, the research community focused more on the field of communication. That same year, the Physics Nobel Prize was awarded for "groundbreaking achievements concerning the transmission of light in fibers for optical communication".

```
1  PREFIX spif: <http://spinrdf.org/spif#>
2  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3  PREFIX jur: <http://sweet.jpl.nasa.gov/2.3/humanJurisdiction.owl#>
4  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
6  PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
7  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
8
9  SELECT ?venue ?year (COUNT(?paper) AS ?numPapers) WHERE {
```

```
10
11      # get the most important venues (the ones with at least 800 papers published)
12      {
13          SELECT ?venue (COUNT(?paper) AS ?totPapers) WHERE {
14              ?paper :publishedIn ?venue.
15          }
16          GROUP BY ?venue
17          HAVING (?totPapers > 800)
18          ORDER BY DESC (?totPapers)
19      }
20
21      # get the number of paper published in the most important venues for each year
22      ?paper :publishedIn ?venue;
23        :hasYear ?year.
24  }
25  GROUP BY ?venue ?year
26  ORDER BY ASC (?year)
```

papers per venue during the years

Query 8 - **papersPerCategory**

(da capire se fare il conteggio totale o per anno e se considerare anche le sottocategorie)

```
1   PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
2   PREFIX : <http://www.semanticweb.org/a3d/ontologies/2024/10/nobelOntology/>
3   # Extracts the number of papers we have for each category over the years --> the most
        studied research areas over the years
4   SELECT ?year ?category (SUM(?howmany) AS ?totalPapers) WHERE {
5       # Inner query to extract the number of papers published in journals that have at
            least one category that is a top concept of our skos scheme
6       {
7         SELECT ?year ?category (COUNT(DISTINCT ?paper) AS ?howmany) WHERE {
8               ?journal :hasJournalCategory ?category .
9                   :journalCategoryScheme skos:hasTopConcept ?category .
10              ?paper :publishedIn ?journal ;
11                  :hasYear ?year .
12          }
13          GROUP BY ?year ?category
14      }
15      UNION
16      # Inner query to extract the number of papers published in journals that have at
            least one category that is a subcategory of a top concept category
17      {
18          SELECT ?year ?category (COUNT(DISTINCT ?paper) AS ?howmany) WHERE {
19              ?journal :hasJournalCategory ?cat .
20                  ?cat skos:broaderTransitive ?category .
21                  ?paper :publishedIn ?journal ;
22                      :hasYear ?year .
23          }
24      GROUP BY ?year ?category
25      }
26  }
27  GROUP BY ?year ?category
28  ORDER BY DESC (?totalPapers)
```