SILURIAN: a Sparql vIsuaLizer for UndeRstanding querIes And federatioNs

Simón Castillo, Guillermo Palma and Maria-Esther Vidal
{scastillo,gpalma,mvidal}@ldc.usb.ve

Motivation

Exploration on the number and size of Linked Datasets
- SPARQL Endpoints to access RDF data.
- Applications from different domains require to gather data from several endpoints.
- Performance of Federated SPARQL queries can be affected by diverse parameters, e.g., number of triple patterns, endpoints and shape of the query.

Approach

Goals:
- Visualize SPARQL queries and federations to understand the complexity of the different plans.

Demonstration Use Cases

Query: Drugs that possibly target Leukemia

```sparql
PREFIX drugbank: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/>
PREFIX dbcategory: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugcategory/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbowl: <http://dbpedia.org/ontology/>
PREFIX kegg: <http://bio2rdf.org/ns/kegg#>
PREFIX diseasesome: <http://www4.wiwiss.fu-berlin.de/diseasome/resource/diseases>

SELECT DISTINCT ?drug1
WHERE {
  ?drug1 drugbank:possibleDiseaseTarget diseasesome:673 .
  ?o drugbank:genbankIdGene ?g .
  ?o drugbank:hprdId ?hp .
  ?o drugbank:swissprotName ?sn .
  OPTIONAL { ?drug owl:sameAs ?drug5 .
    ?enzyme rdf:type kegg:Enzyme .
    ?reaction kegg:xEnzyme ?enzyme .
    ?reaction kegg:equation ?equation .
  }
}
```

Discussion

Triple patterns bound to general predicates, e.g., from RDFS or OWL,
- all endpoints may need to be contacted to produce complete answers.

Queries with large number of triple patterns may
- be decomposed in a large number of sub-queries, and
- require to execute costly sub-queries.

Data partition and replication may negatively impact on performance
- Relevant endpoints increase according to data fragments,
- Execution time may be affected by vertical fragmentation,
- Completeness may be impacted for horizontal fragmentation.

Conclusions and Future Work

Summarizing,
- SILURIAN visualizes parameters that impact on the complexity of federated queries, e.g.,
  - data fragmentation and replication,
  - triple patterns bound to general predicates,
  - query shape, and
  - between answer completeness and execution time.

In the Future,
- SILURIAN will visualize plans produced by state-of-the-art federated engines,
- Extend SILURIAN to understand the impact of dynamicity and data updates.