

Federating and querying heterogeneous and distributed Web APIs and triple stores

Tarcisio Mendes de Farias^{1,2}, Christophe Dessimoz^{1,2}, Aaron Ayllon Benitez³, Chen Yang⁴, Jiao Long⁵ and Ana-Claudia Sima^{1,2}

Summary

Today's international corporations such as BASF, a leading company in the crop protection industry, produce and consume more and more data that are often fragmented and accessible through Web APIs. In addition, part of BASF's data of interest are stored as triples and accessible with the SPARQL query language. Homogenizing the data access modes and the underlying semantics of the data without modifying or replicating the original data sources become important requirements to achieve data integration and interoperability. In this work, we propose a federated data integration architecture within an industrial setup, that relies on an ontology-based data access method. Our performance evaluation in terms of query response time showed that most queries can be answered in under 1 second.

Methods

- **Data integration**
 - Decentralized approach
 - Always up-to-date
 - Reduced maintenance
- **Virtual Knowledge Graphs**
 - Data retrieved directly from their original sources via Web APIs are **combined during query execution**
 - SPARQL micro-services
- **Semantic interoperability**
 - **Ontologies** (e.g. ORTH ontology, UniProt core ontology)

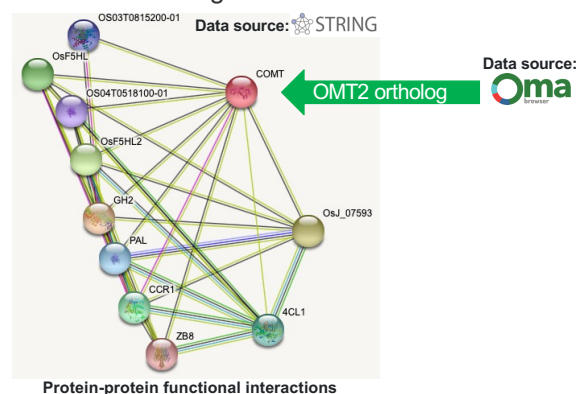
SPARQL endpoints:
OMA: <https://sparql.omabrowser.org/sparql>
StringDB: several SPARQL endpoints, there is one per REST API function

Results and a query example over OMA and STRING public databases

Table 1. Benchmark results of the federated data integration architecture implementation in terms of query execution time and retrieved results. The means and standard deviations related to the execution time of each query are defined in seconds. For more detail, see: <https://purl.org/stringdb/query-eval>

Query	Mean (s)	Std deviation	Retrieved results
Q1	0.95	0.51	10
Q2	0.31	0.03	0
Q3	0.74	0.12	26
Q4	0.26	0.02	3
Q5	0.32	0.01	1
Q6	0.91	0.21	36
Q7	0.37	0.07	0
Q8	1.33	0.48	10

Q8: what are the direct **protein-protein functional interactions** of a rice gene that **is orthologous** to the OMT2 wheat gene?



■ <https://purl.org/stringdb/sparql-git>



Conclusion

- We built a semantically enriched abstract layer with virtual knowledge graphs on top of the Web API functions that access the physical layer (i.e., data sources).
- Federated data integration prototype to interoperate Web APIs with native RDF stores
- Future work: the comparison of SPARQL-MS with other commercial and non-commercial OBDA approaches over Web APIs

Funding and acknowledgements



Contact

tarcisio.mendes@sib.swiss
www.sib.swiss

¹SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
²University of Lausanne, Lausanne, Switzerland
³BASF Digital Solutions SL, Madrid, Spain
⁴BASF, Ghent, Belgium
⁵Ghent University, Ghent, Belgium