

SPARQL Query Formulation and Execution using FedViz

Syeda Sana e Zainab¹, Ali Hasnain¹, Muhammad Saleem², Qaiser Mehmood¹, Durre Zehra¹, and Stefan Decker¹

¹ Insight Centre for Data Analytics, National University of Ireland, Galway
firstname.lastname@insight-centre.org

² Universität Leipzig, IFI/AKSW, PO 100920, D-04009 Leipzig
{lastname}@informatik.uni-leipzig.de

Abstract. Health care and life sciences research heavily relies on the ability to search, discover, formulate and correlate data from distinct sources. Although the Semantic Web and Linked Data technologies help in dealing with data integration problem there remains a barrier adopting these for non-technical research audiences. In this demo we present FedViz, a visual interface for SPARQL query formulation and execution. FedViz is explicitly designed to increase intuitive interaction with data from distributed sources. FedViz facilitates both federated and non-federated SPARQL queries formulation through a visual interface and uses FedX for query execution and the corresponding results retrieval. We evaluate the usability of our system by using the standard system usability scale questionnaire. Our overall usability score of **74.16** suggests that FedViz interface is easy to learn, consistent, and adequate for frequent use.

1 Introduction

The researchers in health care, life sciences and biomedical also known as domain users, adopted Semantic Web and Linked Data technologies due to the data integration challenges faced as a result of excessive data produced [2,8]. The use of these technologies facilitated the domain users for issuing structured SPARQL queries over highly heterogeneous data spread over diverse data sources [1]. Such structured queries are important to drive meaningful biomedical correlations. Meaningful query such as *find out the drugs and their interactions with other drugs that help in curing lung cancer?* can only be answered by querying and aggregating data from more than a single data sources. Due to lack of Semantic Web expertise, writing the desired SPARQL queries to retrieve such data is a challenging task for most of the biologist or clinical researchers [4]. Even for the computer scientist, assembling a federated SPARQL query (i.e., the one which integrates data from multiple data sources) is a time-consuming and technical task as it heavily relies on the familiarity with the data representation and the granularity of each federated data source [6,3,5]. An alternative to this is an intuitive and interactive platform that can facilitate domain users to assemble complex meaningful SPARQL queries (can be SPARQL 1.0 federated queries as well) through visual interface. To this end, we introduce FedViz, an intuitive visual query interface for formulating and running complex SPARQL queries. FedViz allows user to select concepts and properties from

multiple datasets using nodes and edges and makes SPARQL query in a background. Furthermore, it also allows users to edit the resultant SPARQL query before sending it to the query execution engine. The assembled SPARQL query is then executed through FedX [9], a well-known SPARQL query federation engine. Finally, the query execution results are displayed by FedViz. Note FedX can execute both federated (without using SPARQL SERVICE clause) and non-federated SPARQL queries. As an example showcase, in this demo, six real time data sources Kegg (<http://www.genome.jp/kegg/>), Drugbank (<http://www.drugbank.ca/>), DailyMed (<http://dailymed.nlm.nih.gov/dailymed/index.cfm>), Medicare (<http://wifo5-03.informatik.uni-mannheim.de/medicare/>), Sider (<http://wifo5-03.informatik.uni-mannheim.de/sider/>), and Disasome (<http://wifo5-03.informatik.uni-mannheim.de/diseasome/>) are selected to visually construct the SPARQL queries.

2 FedViz

FedViz (Demo available at: <http://srygal86.deri.ie/FedViz/index.html>) is a browser based application that provides users with a flexible visual interface to formulate and execute SPARQL queries. Currently, FedViz visualises life sciences datasets and facilitate users to draw meaningful biological correlations including drug-drug interaction, drug-disease interaction, drug-side effect interaction and disease-gene interaction. FedViz supports full range of SPARQL1.1 constructs either selected through interface or can be directly added to the query in SPARQL Editor.

2.1 Query Formulation

In this section, we demonstrate our visual query formulation and execution process by using an example scenario, i.e., find the drug-disease interaction from the aforementioned data sources.

Scenario: Drug-Disease Interaction: *Drugs with their inchi key, molecular weight average and chemical structure for curing disease Alzheimer and having molecular weight average greater than 200:*

This query requires data integration from Drugbank (containing drugs information) and Disasome (containing disease information) and be formulated by using the following step-by-step approach (ref., Fig. 1):

1. The first step is to identify how Drugbank and Disasome datasets are connected to each other? This connectivity (i.e., via classes `drugbank:drug` and `diseasome:disease`) can be found by using the Class visualisation View of FedViz.
2. User selects Drugbank from the Datasets Selection box (window A).
3. The visualisation for DrugBank dataset can be seen in Window B where he selects `drugbank:drug` class.
4. Properties (i.e., inchi key, molecular weight average, possible disease target, and chemical formula) can be selected from Property Selection box (window D).
5. Step 2-4 are now followed for Disasome dataset, i.e., select `diseasome:disease` class and it's name property (window C).

Fig. 1: Federated Query Execution Step by Step Approach.

6. Next, FedViz SPARQL Query Editor allows user to add constraints to the formulated query such as select projection variables, apply SPARQL LIMIT, FILTER, ORDER BY clauses, and can further edit the query according to his choice (window E and F).
7. The final query can be seen on submission (window G).
8. Query is executed over FedX and the retrieve results are displayed by FedViz (Result window H).
9. Finally, by selecting any URI from the retrieved result, FedViz provided detailed information regarding that entity (Data Exploration window I).

3 Evaluation

To assess the usability of FedViz, we used the standardised, *System Usability Scale* (SUS) [7] questionnaire¹. The SUS is a low-cost, reliable usability scale that can

¹ Our survey can found at: <http://goo.gl/forms/bhReuNgd60>

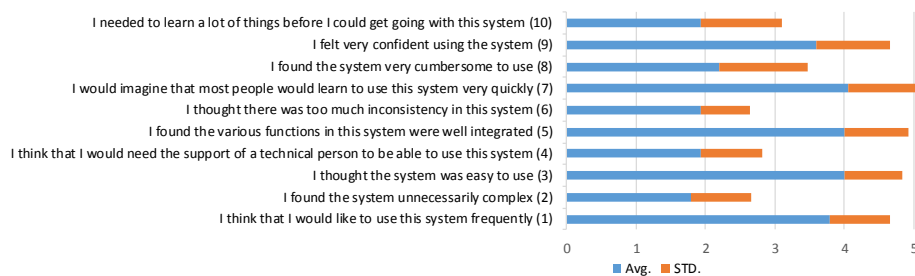


Fig. 2: Result of usability evaluation using SUS questionnaire.

be used for global assessments of systems usability[7]. The survey was filled by 15 users² (by 7th July 2015). We achieved a mean usability score of **74.16** indicating a high level of usability according to the SUS score. The average scores (out of 5) for each survey question along with standard deviation is shown in Figure 2. The responses to question 1 suggests that FedViz is adequate for frequent use (average score to question 1 = 3.8 ± 0.86) by users all of type. The responses to question 7 (average score 4.06 ± 0.96) suggests that most people would learn to use this system very quickly and the responses to question 3 indicates that FedViz is easy to use (average score 4 ± 0.84).

References

1. B. Chen, D. J. Wild, Q. Zhu, Y. Ding, X. Dong, M. Sankaranarayanan, H. Wang, and Y. Sun. Chem2bio2rdf: A linked open data portal for chemical biology. *arXiv:1012.4759*, 2010.
2. H. Chen, T. Yu, and J. Y. Chen. Semantic web meets integrative biology: a survey. *Briefings in bioinformatics*, 14(1):109–125, 2013.
3. A. Hasnain, R. Fox, S. Decker, and H. F. Deus. Cataloguing and linking life sciences LOD Cloud. In *1st International Workshop on Ontology Engineering in a Data-driven World collocated with EKAW12*, 2012.
4. A. Hasnain, M. R. Kamdar, P. Hasapis, D. Zeginis, C. N. Warren Jr, et al. Linked Biomedical Dataspace: Lessons Learned integrating Data for Drug Discovery. In *International Semantic Web Conference (In-Use Track), October 2014*, 2014.
5. A. Hasnain, S. S. E. Zainab, M. R. Kamdar, Q. Mehmood, C. Warren Jr, et al. A roadmap for navigating the life sciences linked open data cloud. In *International Semantic Technology (JIST2014) conference*, 2014.
6. M. R. Kamdar, D. Zeginis, A. Hasnain, S. Decker, and H. F. Deus. Reveald: A user-driven domain-specific interactive search platform for biomedical research. *Journal of biomedical informatics*, 47:112–130, 2014.
7. J. R. Lewis and J. Sauro. The factor structure of the system usability scale. In *HCD*. 2009.
8. M. Saleem, S. S. Padmanabhuni, A.-C. N. Ngomo, A. Iqbal, J. S. Almeida, S. Decker, and H. F. Deus. Topfed: Tcga tailored federated query processing and linking to lod. *Journal of Biomedical Semantics*, 2014.
9. A. Schwarte, P. Haase, K. Hose, R. Schenkel, and M. Schmidt. Fedx: Optimization techniques for federated query processing on linked data. In *ISWC*. 2011.

² Users from AKSW, University of Leipzig and INSIGHT Centre, National University of Ireland, Galway. Summary of the responses can be found at: <https://goo.gl/ZOrJx9>