Linked Data Interfaces for Non-Expert Users

Patrick Hoefler

Know-Center, Graz, Austria
`phoefler@know-center.at`

Abstract. Linked Data has become an essential part of the Semantic Web. A lot of Linked Data is already available in the Linked Open Data cloud, which keeps growing due to an influx of new data from research and open government activities. However, it is still quite difficult to access this wealth of semantically enriched data directly without having in-depth knowledge about SPARQL and related semantic technologies. The presented dissertation explores Linked Data interfaces for non-expert users, especially keyword search as an entry point and tabular interfaces for filtering and exploration. It also looks at the value chain surrounding Linked Data and the possibilities that open up when people without a background in computer science can easily access Linked Data.

Keywords: linked data, interfaces, semantic web, sparql, rdf

1 Motivation

The Linked Open Data cloud provides an impressive wealth of semantically enriched, openly available Linked Data. However, this Linked Data is basically only accessible for experts in semantic technologies who know how to write SPARQL queries. And even for those who know how to use SPARQL, it can be quite laborious at times, especially while trying to explore an unknown SPARQL endpoint.

Therefore, the goal of this dissertation is to research easy-to-use interfaces for accessing Linked Data. These interfaces should be usable without any prior knowledge of SPARQL or other semantic technologies.

2 Research Questions

The central research question is:

- **How can Linked Data interfaces for non-expert users look like?**

  Subsequent research questions are:

- How can current search engine paradigms be used for Linked Data?
- How can Linked Data be displayed, filtered, and explored in tabular form?
- How can value be created when non-expert users can access Linked Data?

3   Approach and Research Methodology

The first phase of this dissertation dealt with some of the fundamental problems related to the question of easy access. In [1], we looked into ways to turn search keywords into URIs and applied the technique in a simplified end-user interface for accessing Linked Data. This approach was extended into CAF-SIAL, a proof-of-concept application described in [2] and [3]. CAF-SIAL helped users to search information about concepts in the Linked Open Data cloud without having to know any of the mechanics of the Semantic Web.

We also created a Linked Data value chain model [4] that conceptualized the actors and processes in the Linked Data ecosystem. Though we did not pursue that strand of research further at the time, it turned out to become much more relevant to this dissertation later on.

The second phase of this dissertation is closely related to the CODE project [5], a research project funded by the European Union. As described in [6], the vision of CODE is to establish a sophisticated ecosystem for Linked Data. The current focus of this dissertation is the Linked Data Query Wizard[1] as part of CODE's Visual Analytics work package [7]. The goal of the work package is to develop a web-based visual analytics platform that enables non-expert users to engage in a visually supported, collaborative analysis of Linked Data, and the Linked Data Query Wizard will play a crucial role in this undertaking.

The current research methodology follows the principles of agile development and rapid prototyping: Small implementation cycles, resulting in new versions of the prototype on a weekly, sometimes daily basis.

4   The Linked Data Query Wizard

The working hypothesis for the Linked Data Query Wizard is: There's not a lot of people who speak SPARQL and are familiar with graph structures. On the other hand, many people know spreadsheet applications like Microsoft Excel. Therefore, the idea is to develop a web-based tool that brings the graph structure of Linked Data into tabular form (see figure 1) and provides easy-to-use interaction possibilities for filtering and exploring Linked Data by using metaphors and techniques the users already know.

4.1   Related Work

Although the Semantic Web has matured in recent years, and semantic technologies have become quite powerful, the Linked Open Data cloud is still only accessible for semantic technology experts and programmers. The problem of easy-to-use interfaces for accessing Linked Data is still largely unsolved. The majority of current tools are not aimed at non-expert users. As an example, the popular Semantic Web search engine Sindice [8] is practically unusable for people without a deep understanding of semantic technologies.

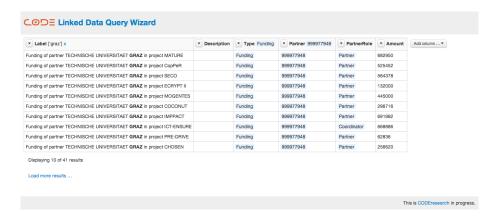[1] `http://code.know-center.tugraz.at/search`

Fig. 1. Early beta version of the Linked Data Query Wizard.

Currently, only very few web-based tools use tables for representing Linked Data. One such example would be Freebase Parallax [9]. Although its main feature is the ability to browse sets of related things, it also provides a table view for these result sets. Another online tool that shares similarities with our prototype is the Falcons Explorer [10]. Both tools feature a search box as the main entry point – an idea that is also central to our prototype. However, in both tools, the table view is not the central focus.

Another tool that shares a few similarities with our prototype is OpenRefine [11] (formerly known as Google Refine and Freebase Gridworks). It supports RDF, and there are also extensions such as LODRefine [12] that focus on Linked Data – however, OpenRefine's main focus is cleaning up tabular data, and it's also not available as a web service, even though its main interface is browser-based.

Our prototype also supports the current working draft of W3C's RDF Data Cube Vocabulary [13] which provides a semantic framework for expressing statistical data sets as Linked Data. Data sets that comply with the RDF Data Cube standard can easily be displayed, filtered, and explored using the Linked Data Query Wizard. To the best of our knowledge, there are currently no other tools with a feature set similar to our prototype that support RDF Data Cubes.

4.2   Limitations

Currently, the biggest limitation of the prototype is the rather demanding set of requirements it imposes on the SPARQL endpoints that serve as its back end. One critical feature that is needed for our current approach is support for full-text search. Sadly, full-text search is sorely lacking from the current SPARQL specification, which is why certain SPARQL endpoints have come up with workaround solutions. Therefore, only Virtuoso and bigdata are currently supported as SPARQL endpoints by the Linked Data Query Wizard.

The Linked Data Query Wizard also makes use of certain SPARQL 1.1 features, especially the aggregation functions. The `COUNT()` function is critical and already in use by the current prototype for displaying the number of results for a given query.

4.3   Evaluation

For evaluating the prototype, both formative as well as summative evaluations are planned:

The formative evaluations have already started informally, mostly with project team members, and with increasing feature completeness will also include friendly users outside of the team with a decreasing degree of expertise. The last formative evaluations will include non-expert users.

There are also summative evaluations planned. Some of them will be part of challenges and the planned use of the prototype in university courses, but additional quantitative and qualitative user studies might be conducted as needed.

4.4   Initial Results

The Linked Data Query Wizard is currently available online as an early beta version. In its current form, it offers two entry points: Users can either initiate a keyword search, or they can select any available dataset, represented as an RDF Data Cube. In both cases, the users get presented with the results in tabular form, similar to what they are used from spreadsheet applications. They can choose which columns (i.e. RDF predicates) they are interested in, and they can set filters to narrow down the displayed data.

Though the current functionality of the prototype is still rather limited, first usage experiments have shown that the tool can be helpful in exploring the data and respective data structures of unknown SPARQL endpoints.

5   Conclusions and Future Work

The creation process of this dissertation can be divided into two phases: The early first phase, focusing on keyword search for Linked Data and the value chain of the Linked Data ecosystem, and the current second phase, focusing on the Linked Data Query Wizard, a novel approach for filtering and exploring Linked Data.

The next steps regarding the prototype will be to expand its functionality, focusing on better filter mechanisms as well as more advanced exploration features that incorporate the underlying semantic structure. Additionally, the Linked Data Query Wizard will be integrated with a tool for visualizing Linked Data as well as a semantic enrichment service for turning generic RDF into RDF Data Cubes.

The development of the prototype will continue throughout the rest of the year, leading to a final evaluation at the beginning of 2014.

The writing process of the dissertation has already begun and will intensify as the year progresses. The majority of the dissertation should be finished by the end of 2013, leaving only the final evaluation results as well as the finishing touches for the beginning of 2014.

The dissertation should be finished in the first half of 2014.

References

1. Latif, A., Afzal, M. T., Hoefler, P., Us Saeed, A., Tochtermann, K.: Turning Keywords into URIs: Simplified User Interfaces for Exploring Linked Data. In: Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 76-81. ACM, New York (2009)
2. Latif, A., Afzal, M. T., Us Saeed, A., Hoefler, P., Tochtermann, K.: CAF-SIAL: Concept Aggregation Framework for Structuring Informational Aspects of Linked Open Data. In: Proceedings of NDT 2009. Ostrava (2009)
3. Latif, A., Us Saeed, A., Hoefler, P., Tochtermann, K., Afzal, M. T.: Harvesting Pertinent Resources from Linked Open Data. In: Journal of Digital Information Management, 8(3), pp. 205–212. Digital Information Research Foundation, Chennai (2010)
4. Latif, A., Hoefler, P., Stocker, A., Us Saeed, A., Wagner, C.: The Linked Data Value Chain: A Lightweight Model for Business Engineers. In: Proceedings of I-KNOW '09 and I-SEMANTICS '09. Verlag der Technischen Universität Graz (2009)
5. CODE: Commercially Empowered Linked Open Data Ecosystems in Research, `http://code-research.eu/`
6. Stegmaier, F., Seifert, C., Kern, R., Hoefler, P., Bayerl, S., Granitzer, M., Kosch, H., et al.: Unleashing Semantics of Research Data. In: Proceedings of the Second Workshop on Big Data Benchmarking (WBDB2012.in). Pune, India (2012)
7. CODE: Visual Analytics, `http://code-research.eu/visual-analytics`
8. Tummarello, G., Delbru, R., Oren, E.: Sindice.com: Weaving the open linked data. The Semantic Web, 552-565. Springer Berlin Heidelberg (2007)
9. Huynh, D. F., Karger, D. R.: Parallax and Companion: Set-based Browsing for the Data Web. WWW Conference, ACM, 2009.
10. Cheng, G., Wu, H., Gong, S., Ge, W., Qu, Y.: Falcons Explorer: Tabular and Relational End-user Programming for the Web of Data. Semantic Web Challenge, 2010.
11. OpenRefine, `https://github.com/OpenRefine`
12. LODRefine, `http://code.zemanta.com/sparkica/`
13. W3C: RDF Data Cube Vocabulary, `http://www.w3.org/TR/vocab-data-cube/`