ExConQuer Framework - Softening RDF Data to Enhance Linked Data Reuse

Judie Attard, Fabrizio Orlandi, Sören Auer

University of Bonn
attard@iai.uni-bonn.de, orlandi@iai.uni-bonn.de, auer@cs.uni-bonn.de

Abstract. Efforts towards the wider adoption of Linked Open Data, including the implementation of Linked Data principles and the consumption of Linked Data, are evident in the existing literature and available tools. Yet, these efforts rarely cater for stakeholders who are not familiar with RDF or SPARQL. Hence, we propose the ExConQuer Framework, which facilitates the publication and consumption of RDF in a variety of generic formats. In this manner, any stakeholder can export and work with RDF data in the formats they are most accustomed with, thus lowering the entry barrier to the use of semantic technologies, and possibly enabling the exploitation of Linked Open Data to its full potential.

Keywords: linked open data, open data consumption, open data publication, RDF softening

1 Introduction

The extraordinary growth in volume of the Linked Open Data (LOD) Cloud[1] is evidence enough that Linked Data practices are being adopted at an increasing rate, and accessibility to raw data, especially in recent years, is being given considerable importance [1]. Whereas raw data used to be published in barely-interpretable formats such as PDF or CSV, the implementation of Linked Data practices has achieved a more meaningful representation of the same data on the Web. Yet, although barriers to information access have been lowered through various means, it does not mean that the average stakeholder can easily locate, access, or most importantly reuse such data. Individuals wanting to reuse Linked Data might be more acquainted with file formats such as JSON, XML or CSV, and are not necessarily familiar with RDF, SPARQL, or the datasets' underlying schema, which can be perceived to be too complex to learn. Stakeholders might therefore end up either forgoing any attempt to reuse Linked Open Data, or otherwise reuse it without truly exploiting it to its full potential, for example by downloading a data dump rather than specifically querying and reusing the required data. Unfortunately, the emergence of a wide number of tools supporting people to publish their data as Linked Open Data[2] has not been complemented

---

[1] http://lod-cloud.net/
[2] http://www.w3.org/wiki/LinkedData

by approaches supporting them to consume existing Linked Data in formats other than RDF [1]. While Linked Data publishing tools are useful in order to ensure the best quality data is published, it is of no use if the consumers do not have the tools or the expertise to exploit it.

In this paper we propose the ExConQuer (Extract, Convert and Query) Framework[3]. This encompasses a number of tools and technologies intended for less experienced or novice users of Linked Data. Through *RDF softening* (as opposed to semantic lifting) we generate semantically-shallow representation formalisms of RDF data views, whilst retaining the semantic richness of RDF through provenance information. Our aim is therefore to lower the entry barrier to Linked Data reuse, and enable stakeholders to exploit the full potential of Linked Data without requiring to know RDF or SPARQL.

2   Approach

The ExConQuer Framework aims to assist stakeholders in consuming and publishing Linked Data. We enable them to explore existing datasets, construct SPARQL queries, generate different views of the results, and publish a representation of the undertaken tasks to enable further use. These functions are provided through the **Query Builder Tool**[4], the **RDF2Any API**, the **ConQuer Ontology**[5], and the **PAM Tool**[6].

Figure 1 shows an overview of the architecture within the framework. The user can explore datasets and create a SPARQL query through the Query Builder Tool, then query a datastore (through a SPARQL endpoint) through API calls. Information pertinent to the executed processes is then persisted in a triple store as *Linked Data Publications*. Linked Data Publications are automatically published in the PAM Tool once a user downloads the results. Consequently, users can access all the Linked Data Publications through the Provenance-Aware Management Tool (PAM Tool), which allows a user to re-run existing queries, or modify them through the Query Builder.

2.1   Query Builder Tool

Our approach is intended to be particularly user friendly and simple, to allow non-experts to easily use the tool to achieve the goal of re-using open data. Through the RDF2Any RESTful API, the Query Builder Tool enables users to navigate through classes, subclasses, instances, and properties without requiring to know the underlying structure of RDF data. The Query Builder Tool allows users to explore open datasets and, through the execution of a few simple steps, generate a SELECT SPARQL query without requiring any prior knowledge of the query language. The UI therefore enables the user to select a dataset, a class

---

[3] More information on the framework, including source code, can be found here: `http://eis.iai.uni-bonn.de/Projects/ExConQuer`

[4] `http://purl.org/net/exconquer/builder`

[5] `http://purl.org/eis/vocab/cqo`
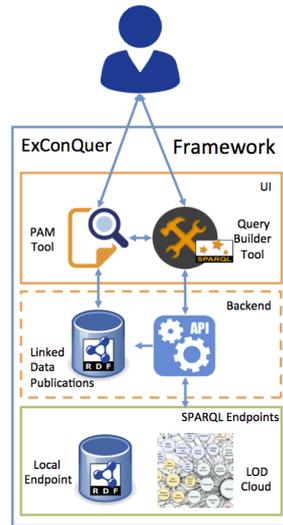
[6] `http://purl.org/net/exconquer`
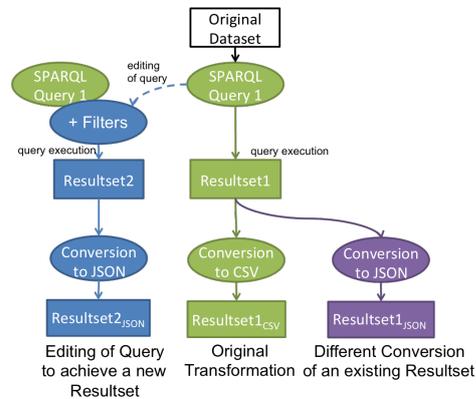
Fig. 1: ExConQuer Framework Architecture

Fig. 2: Possible data flows through ExConQuer Framework

(or a subclass), and properties to include in the results. The user can also add filters to restrict the results even further. After each user selection, the SPARQL query that is generated on the fly is updated, so the user also has the option to modify it. Finally, the user can preview a sample of the results, then download the complete result set in one of the provided formats, namely RDF, CSV, JSON, RDB, and a configurable conversion that is intended for more experienced users. The latter allows a user to convert RDF into potentially any output format, such as XML, KML, TSV (tab separated values) through passing the required parameters in a template.

2.2 RDF2Any API

This RESTful API provides the functionality to the Query Builder Tool. The aim of this API is to hide the complexity of RDF and the datasets' underlying schema. The available API calls include getting all classes in a given dataset, getting classes that match a given keyword (in their labels or URIs), getting the subclasses, instances, and properties of a given class, and converting (softening) a result set in a number of given formats.

2.3 ConQuer Ontology

All the processes executed through the ExConQuer Framework generate what we call a *Linked Data Publication*. A Linked Data Publication consists of all the generated information, including the SPARQL query used, its description, the dataset(s) queried, the initial and target data formats, and the user generating the Linked Data Publication instance. We represent all this data using the

ConQuer Ontology, which reuses concepts from the SPIN vocabulary[7] and the PROV-O Ontology[8]. The use of PROV-O enables us to represent provenance information. The ConQuer Ontology allows us to replicate the resulting Linked Data Publications and edit them to achieve different results (Figure 2). This allows us to implement *RDF softening* without actually compromising on the the richness of RDF representation, as any resultsets in formats other than RDF are linked back to the original data in RDF.

2.4   PAM Tool

We implemented the *PAM Tool* as a provenance-aware publishing and consumption management tool that enables the exploration of Linked Data Publications through a faceted browser. Through the use of the ConQuer ontology, the Linked Data Publications have queryable metadata that enables users to search for specific instances using various criteria, such as by the datasets used and the classes queried for. This tool thus allows users to share, explore, directly edit (through the Query Builder or otherwise), and re-use Linked Data Publications, whilst keeping data lineage intact.

3   Conclusion

The ExConQuer Framework provides solutions that encourage and enable the re-use of open data. Our approach is targeted towards inexperienced users, thus we do not assume that users of our framework are familiar with the Linked Data paradigm. A preliminary evaluation indicated that the tools are intuitively easy to use and useful with regards to their intended use. Through the Query Builder Tool, we aim to lower the entry barrier for any stakeholder requiring the use of Linked Open Data. We enable the user to explore existing Linked Data and generate a SPARQL query, then proceed to download and convert the results in a number of formats. Through the PAM Tool, the user is able to explore existing queries executed on various datasets through filters, and re-load them on the Query Builder tool to edit or re-run them. We showcase all these functionalities in the Demo Video[9] (replicable in `http://purl.org/net/exconquer/builder`).

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. (2009)
2. Langegger, A., Woss, W.: Rdfstats - an extensible rdf statistics generator and library. 2012 23rd International Workshop on Database and Expert Systems Applications.
3. Rozell, E., Erickson, J.S., Hendler, J.: From international open government dataset search to discovery: a semantic web service approach. In: ICEGOV. (2012)

---

[7] `http://www.w3.org/Submission/spin-overview/`

[8] `http://www.w3.org/TR/prov-o/`

[9] `https://youtu.be/ZqS1dOiGcss`