

Exploring semantic datasets with RDF Surveyor

Guillermo Vega-Gorgojo, Martin Giese, Laura Slaughter

Department of Informatics, University of Oslo, Norway

Abstract. Despite the increasing availability of RDF data, exploring and browsing semantic datasets is still a daunting task for lay users. RDF Surveyor is an easy-to-use visual exploration tool that can be plugged in any CORS-enabled SPARQL 1.1 endpoint without requiring any installation. RDF Surveyor offers an intuitive user interface that gives an overview of the repository contents, supports class navigation, and individual visualisation. RDF Surveyor is very lightweight and can handle large datasets such as DBpedia in a responsive way.

1 Introduction

With the advent of RDF data across all domains, there is a need for exploring these datasets not only by Semantic Web experts, but also by lay users. However, [2] reports that even the former group of users are severely limited in their ability to obtain a good understanding of the structure and content of semantic datasets, especially if they are large. While some repositories provide HTML representations of the contained resources, e.g. http://dbpedia.org/page/The_Surrender_of_Breda, this is not sufficient to grasp the structure of a dataset. As an alternative, a SPARQL editor can be used to explore a dataset, but this requires proficiency in SPARQL and the process can be rather tedious if the user is not familiar with the data content and structure.

Enabling lay users to explore an RDF dataset requires a tool that provides an effective visual interface for analysing the contents, giving an overview of the dataset, and supporting class navigation. Moreover, such a tool should be completely generic and not require any installation to facilitate its usage. Finally, it should work even with large datasets containing billions of triples. To the best of our knowledge, there is no tool that meets all the requirements above. As an example, LodView¹ is a prototypical semantic browser that presents RDF data in a tabular form and allows link traversal, but it does not provide any overview of the dataset and class navigation is not supported. Visual query tools like PepeSearch [4], OptiqueVQS [3] and SemFacet [1] support faceted search over RDF datasets and can be used for exploration purposes. PepeSearch and OptiqueVQS present a flat list of the classes in the repository as a starting point, while SemFacet asks for a set of keywords as input in order to build a faceted interface. However, all these tools require a non-trivial installation and are not able to cope with large datasets such as DBpedia.

¹ <http://lodview.it/>

In this demonstration we present RDF Surveyor, a novel and lightweight tool for exploring semantic datasets. It offers an easy-to-use interface that completely hides the RDF and SPARQL syntax from the user, providing an overview of the repository contents, intuitive class navigation, and visualisation of individuals. Notably, RDF Surveyor does not require any installation and can handle large datasets.

2 Overview of RDF Surveyor

RDF Surveyor is a Javascript tool available under an Apache 2 license on GitHub.² A live version of RDF Surveyor is also available at <http://tools.sirius-labs.no/rdfsurveyor> for testing purposes. The tool runs entirely in a browser without relying on any server component, while target repositories only require a CORS-enabled³ SPARQL 1.1 endpoint.

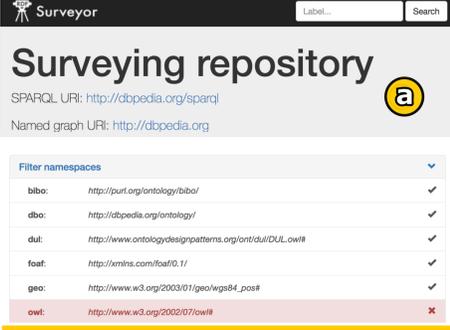
RDF Surveyor solely relies on the SPARQL endpoint to build the user interface, submitting queries on demand based on the actions performed by the user, and caching results in the session storage of the browser to minimise the number of queries posed. To begin the exploration of a dataset, the user only has to enter the URI of the SPARQL endpoint and optionally the URI of a named graph in a web form. RDF Surveyor then submits a bootstrapping query to get the upper classes in the dataset; when results arrive, further queries are sent to obtain labels, comments, direct subclasses, and number of individuals for the upper classes. All this information is used to create an overview page of the repository with the employed namespaces (Fig. 1(a)) and the list of upper classes sorted by the number of direct subclasses and members (Fig. 1(b)).

The user can then filter out unwanted namespaces, drill down the subclasses of a listed class, or select a class. In the latter case, RDF Surveyor prepares a page with the class URI, label, comment, superclasses, subclasses, and member individuals (Fig. 1(c)) – new queries are sent in case of missing information. The user can navigate the class taxonomy, and there are controls for paginated browsing and searching individuals by label – supported by limit/offset modifiers and regex expressions. If an individual is selected, a page is created with the individual URI, label, comment, class types, datatype properties, and direct and inverse object properties (Fig. 1(d)). RDF Surveyor also tries to find a picture URL and geographic coordinates to feed an image and a map widget, respectively. The user can jump to other individual or class pages by clicking the corresponding links. In every page, there is a search textbox for finding classes and individuals by label in the whole dataset (see the navbar in Fig. 1(a)).

Other features of RDF Surveyor include its RESTful design, so all pages expose consistent URIs that can be safely shared and bookmarked; support for multilingual datasets, taking the browser language as first option and English in other case; and, finally, the user interface is adapted to mobile devices.

² <https://github.com/guiveg/rdfsurveyor>

³ <https://enable-cors.org/>



Surveying repository

SPARQL URI: <http://dbpedia.org/sparql>

Named graph URI: <http://dbpedia.org>

Filter namespaces

| | | |
|-------|-----------------------------------------------------------------------------------------------------------------------------|---|
| bibo: | http://purl.org/ontology/bibo/ | ✓ |
| dbo: | http://dbpedia.org/ontology/ | ✓ |
| dul: | http://www.ontologydesignpatterns.org/ont/dul/DUL.owl# | ✓ |
| foaf: | http://xmlns.com/foaf/0.1/ | ✓ |
| geo: | http://www.w3.org/2003/01/geo/wgs84_pos# | ✓ |
| owl: | http://www.w3.org/2002/07/owl# | ✗ |

Upper classes

| | |
|----------------------------|-----------|
| dbo:Place | 18 D +1K1 |
| dbo:Topical concept | 18 D +1K1 |
| dbo:Work | 18 D +1K1 |
| dbo:Written work | 18 D +1K1 |
| dbo:Musical work | 8 D +1K1 |
| dbo:Document | 8 D +1K1 |
| dbo:Artwork | 2 D +1K1 |
| dbo:Cartoon | 2 D +1K1 |
| Show more | |
| dbo:Mean of transportation | 14 D +1K1 |

The Surrender of Breda

http://dbpedia.org/resource/The_Surrender_of_Breda

La rendición de Breda (English: The Surrender of Breda, also known as El cuadro de las lanzas or Las lanzas) is a painting by the Spanish Golden Age painter Diego Velázquez. It was completed during the years 1634–35, inspired by Velázquez's visit to Italy with Ambrogio Spinola, the Genoese general who conquered Breda on June 5, 1625. It is considered one of Velázquez's best works. Jan Morris has called it "one of the most Spanish of all pictures".



Types

dbo:Artwork schema:CreativeWork dbo:Work

Literals

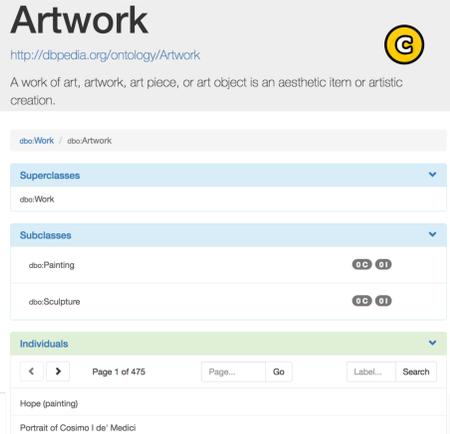
| Property | Value |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| dbo:has abstract | La rendición de Breda (English: The Surrender of Breda, also known as El cuadro de las lanzas or Las lanzas) is a painting by the Spanish Golden Age painter Diego Velázquez. It was completed during the years 1634–35, inspired by Velázquez's visit to Italy with Ambrogio Spinola, the Genoese general who conquered Breda on June 5, 1625. It is considered one of Velázquez's best works. Jan Morris has called it "one of the most Spanish of all pictures". |
| foaf:name | The Surrender of Breda |
| dbo:title | The Surrender of Breda |
| dbo:Wikipedia page ID | 14237621 |
| dbo:Wikipedia revision ID | 666714018 |

Outgoing links

| Property | Target |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| dbo:Artist | Diego Velázquez |
| dbo:author | Diego Velázquez |
| dct:subject | 1634 paintings 1635 paintings Horses in art Velázquez paintings in the Museo del Prado War paintings |
| dbo:thumbnail | http://en.wikipedia.org/wiki/Special:FilePath/Velazquez-The_Surrender_of_Breda.jpg?width=300 |
| prov:wasDerivedFrom | http://en.wikipedia.org/wiki/The_Surrender_of_Breda?oldid=666714018 |

Incoming links

| is Property of | Source |
|------------------------------|--------------------|
| is dbo:Wikipedia redirect of | Surrender of Breda |



Artwork

<http://dbpedia.org/ontology/Artwork>

A work of art, artwork, art piece, or art object is an aesthetic item or artistic creation.

dbo:Work / dbo:Artwork

Superclasses

dbo:Work

Subclasses

dbo:Painting 8 D 01

dbo:Sculpture 8 D 01

Individuals

Hope (painting)

Portrait of Cosimo I de' Medici

Fig. 1: Sample snapshots of RDF Surveyor: (a) Excerpt of the namespaces in DBpedia for filtering. (b) Overview of the upper classes. (c) Description of the **Artwork** class. (d) Description of **The Surrender of Breda** individual.

Table 1: Sample walkthrough with RDF Surveyor.

| ID | Step description | # of queries | Elapsed time (ms) |
|----|---------------------------------------------------------------------|--------------|-------------------|
| S1 | Load DBpedia repository | 53 | 830 |
| S2 | Filter out the owl namespace | 103 | 984 |
| S3 | Expand subclasses of <code>dbo:Work</code> | 19 | 182 |
| S4 | Load the <code>Artwork</code> class | 11 | 131 |
| S5 | Find members of <code>Artwork</code> with <i>breda</i> in the label | 4 | 265 |
| S6 | Load <code>The Surrender of Breda</code> individual | 9 | 507 |

3 Demonstration overview

The demonstration will showcase the exploration of DBpedia or any other dataset exposing a SPARQL endpoint with RDF Surveyor. Table 1 presents a sample walkthrough for producing Fig. 1 – the number of queries submitted and the elapsed time are reported, as measured with a 2014 MacBook Pro. Every step took less than 1 second to complete, so the user experience is not compromised.⁴ Since query results are cached locally, a repetition of the walkthrough does not involve new queries and every step takes then less than 0.1 second.

In Step S1, information about the upper classes is obtained (see Section 2) – note that for every class found, RDF Surveyor asks if it has more than one thousand members, and only requests the actual count in case of a negative answer for performance reasons. S2 is the slowest step due to an expensive query with a `STRSTARTS` function for detecting subclasses of upper classes beginning with the `owl` namespace, finding 49 subclasses such as `dbo:Place` or `dbo:Work`; subsequent queries compiled information about the classes found.

Acknowledgements

This work has been funded by the Norwegian Research Council through BIGMED (IKT 259055), HealthInsight (NFR 247784/O70), and SIRIUS (NFR 237898).

References

1. Arenas, M., Grau, B.C., Kharlamov, E., Marciuška, Š., Zheleznyakov, D.: Faceted search over RDF-based knowledge graphs. *Web Semantics: Science, Services and Agents on the World Wide Web* 37, 55–74 (2016)
2. Dadzie, A.S., Rowe, M.: Approaches to visualising Linked Data: A survey. *Semantic Web* 2(2), 89–124 (2011)
3. Soyly, A., Giese, M., Jimenez-Ruiz, E., Vega-Gorgojo, G., Horrocks, I.: Experiencing OptiqueVQS: a multi-paradigm and ontology-based visual query system for end users. *Universal Access in the Information Society* 15(1), 129–152 (2016)
4. Vega-Gorgojo, G., Giese, M., Heggstøyl, S., Soyly, A., Waaler, A.: PepeSearch: Semantic data for the masses. *PLoS ONE* 11(3), e0151573 (2016)

⁴ <https://www.nngroup.com/articles/response-times-3-important-limits/>