vocab-express - Exploring the Vocabulary of a Dataset

Boris Villazón-Terrazas[1] and Michael Hausenblas[2]

[1] OEG-DIA, FI, Universidad Politécnica de Madrid, Spain
bvillazon@fi.upm.es
[2] Digital Enterprise Research Centre, NUI Galway, Ireland
michael.hausenblas@deri.ie

Abstract. One of the drawbacks of the reuse, review or query LOD datasets is the lack of mechanisms that provides an overview of the structure of them. One step forward to have such overview of the structure is getting the vocabulary used by the datasets. In this paper we present a brief description of vocab-express that aims at exploring the vocabulary of a dataset.

Key words: Linked Data, RDF Dataset, Vocabulary

1 Introduction and Motivation

So far, Linked Data principles and practices are being adopted by an increasing number of data providers, getting as a result a global data space on the Web containing thousands of LOD datasets [**?**]. In order to reuse, review or query a dataset published on the Web of Data, it is important to know the structure of the data. One step forward for knowing in depth the structure of the data is to explore the vocabulary that the dataset is using, and how the dataset is using such vocabulary.

There are available works such as (1) *LODStats*[3] that provides the information related with the vocabulary of given dataset, and (2) *make-void* [4] that computes statistics about RDF files. However, LODStats is thought for the whole set of LOD datasets registered in The Data Hub [5], and it is based on declarative descriptions of those datasets, and *make-void* is thought for RDF files but not for RDF datasets.

In this paper we present vocab-express[6], a simple tool for exploring the vocabulary of a given dataset. The tool provides all the related information to the vocabulary of the dataset (1) list of all classes, (2) list of all the properties, (3) the number of instances of each class, (4) the number instances of each property, (5) the language of labels and comments, of the vocabulary elements.

[3] http://stats.lod2.eu/
[4] https://github.com/cygri/make-void
[5] http://thedatahub.com
[6] http://vocab-express.nodester.com/

2 vocab-express overview

vocab-express is being implemented in node.js[7], which is a platform built on V8[8] for easily building fast, scalable network applications. Figure 2 depicts the workflow of vocab-express. Next we describe the steps of the workflow

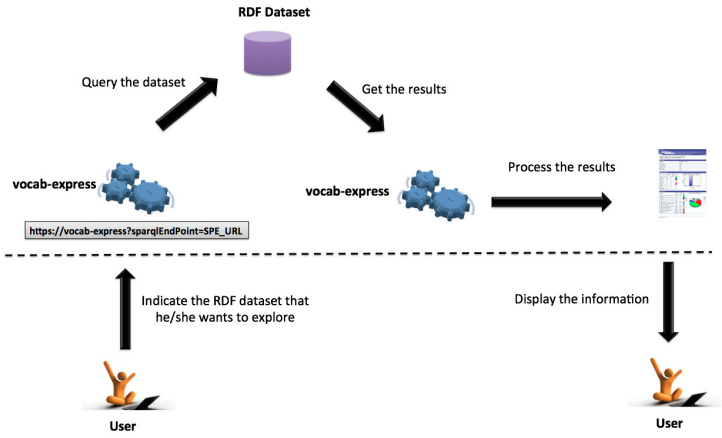

Fig. 1. vocab-express workflow

1. The user indicates the RDF dataset he/she wants to explore by providing its SPARQL endpoint URL
2. vocab-express receives and validates the SPARQL endpoint URL
3. vocab-express queries the SPARQL endpoint and gets results in JSON[9] format
4. vocab-express processes the results and displays the information to the user

The workflow is similar if the user provides the URL of the VoID file instead of the SPARQL endpoint.

3 Conclusions

In this paper we have presented our ongoing work for exploring the vocabulary of an RDF dataset. The tool provides a clear picture of the vocabulary used to model the resources of the dataset and how it is actually used.

As future work we want to evaluate the performance of the tool comparing with approaches not implemented in nodejs. Moreover, we will improve how the information is displayed to the user.

[7] `http://nodejs.org/`
[8] V8 is Google's open source JavaScript engine.
[9] `http://www.json.org`

References

1. T. Heath and C. Bizer. Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web Theory and Technology*, 1(1):1–136, 2011.