

Designing a Step-by-Step User Interface for Finding Provenance Information over Linked Data

Enayat Rajabi¹ and Mohsen Kahani²

¹ Department of Engineering, Saveh Branch, Islamic Azad University, Saveh, Iran
erajabi@iau-saveh.ac.ir

² Computer Engineering Department, Ferdowsi University Of Mashhad, Mashhad, Iran
kahani@um.ac.ir

Abstract. The proliferation of the use of Linked Data, and growth of Linked Open Data (LOD) cloud provide a good environment for interrelating previously isolated datasets. To encourage non-professional users to publish and find their required data easily, a good user interface is needed. Also, as users want to reach trustworthy or more up-to-date information in Linked Data, they would like to have access to the provenance data, as well. In this paper, a new method is presented that not only offers an easy interface for searching data in LOD cloud, but also provides provenance information of data.

Keywords: Linked Data, User-interface, Provenance, Metadata, LOD.

1 Introduction

With regarding to the growing availability of Linked Data on the Web [1], users are looking for approaches that do not limit them dealing with specific tasks.. They shouldn't be also limited to use specific data. Thus, the purpose of this paper is to help users to find their desired data and their provenance in an effective and easy way. Current Linked Data user interfaces such as search engines and browsers are discussed and analyzed in this research. In fact, currently, exploring the Linked Data cloud is difficult for users who are unfamiliar with semantic web concepts such as SPARQL and RDF [2]. Using easy graphical user interfaces for Linked Data, users can search effectively and find their exact data among Linked Datasets. Next, provenance information over Linked Data is considered and a metadata store is proposed for designing a step-by-step user interface and searching valid Linked Data. Finally, this approach is tested using a simple case study and simple metadata store.

2 Linked Data User-Interfaces

Linked Data user interfaces are various and can be classified into different kinds of user interfaces such as search engine, Linked Data browser, mash-ups, and triple query builder, etc. For example, Tabulator [3] browses RDF data on the Linked Data and uses an algorithm that traverses RDF data links and gives a dereferenceable URI. Fenfire [4] browser explores Linked Data with a graph view and highlights the use

cases of exploring and demonstrating Semantic Web data. DBpedia Faceted search ¹ allows users to ask questions and then filters the results.

In most user interfaces for Linked Data, searching or exploring the data is performed similar to that of traditional search engines. Although it is possible to show users more accurate results, they should be able to search and filter the results again to reach the desired data. Users should be allowed to search easily and find the desired data without knowing the Linked Data principles. In some above mentioned user interfaces, users should be familiar with RDF triples. We found some types of interfaces such as Facets and query builders, as more effective ways for exploring data. Facets present filtered options in which end users can filter the results based on some related and limited fields. Because of using Facets and filtering the results in this kind of user interface, our approach is somehow similar to the Facets, but we use a Top-Down filtering method to reach the results, not filtering after showing the results (bottom-up filtering). It means that like advanced options in a search engine, users are guided to reach the results step-by-step by filtering datasets.

3 Finding Data Provenance

Veracity and quality of Linked Datasets are the most important things that should be determined in linked open data cloud regarding to openness of web data. Users can discern acceptance of the data by knowing about detailed history of them. Provenance information can be stored with other metadata or simply by itself [5]. In maintaining provenance, we should consider if it is immutable, or if it can be updated to reflect the current state of its ancestors. If provenance depends on users manually adding annotations instead of automatically collecting it, the burden on the user may prevent complete provenance from being recorded and available in a machine accessible form that has semantic value [6]. Users can also search for datasets based on their provenance metadata, such as locating all datasets that generated by executing a certain workflow. In order to use provenance, a system should allow rich and diverse means to access it.

4 The Proposed Approach

In the proposed approach, users can reach the desired data step-by-step from the metadata store to specific dataset(s) via search form as Figure 1 depicts. The metadata of datasets consist of dataset name, number of sub datasets, dataset ID, dataset usages, its domain and etc. Also details of dataset's relationships with other datasets, such as relation type can be stored in the data store. This helps us to know how each dataset is related with others and also helps users to find their desired data in an easy way.

An RDF data store is defined to store information about datasets and all datasets are encouraged to be registered in this metadata store. A registration process for checking the validity of data is required, as valid information about datasets should be presented to the users when they want to reach to the provenance information of data. After a while, a small cloud of LOD is created.

¹ <http://dbpedia.neofonie.de/browse>

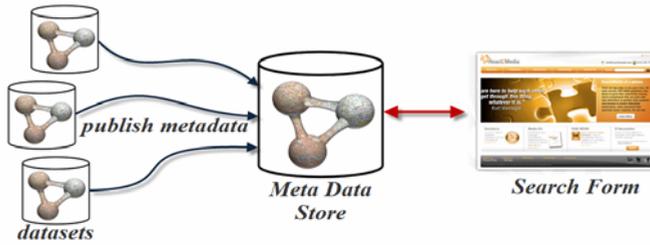


Fig. 1. Storing RDF Metadata in the Data Store

On the other hand, searching for content in a search engine basically comprises some steps like running a search, viewing the list of search results, selecting and viewing a document from the search results list and navigating the selected document. Each search engine usually has a search form that has a simple user interface. In a web search engine, one can restrict her searches by advanced options that are useful when she knows exactly what she wants. In the proposed approach, a search form in which users can select their keywords in a datasets list has been designed. The basic search is started by finding the keyword among the first level datasets in the metadata store in LOD. After finding the data, based on related datasets, users are guided with more information about it. Based on the selected keyword, the related datasets are retrieved from the data store. Each field in search form may depend on a selected field in the previous step in search page and the results of search will be filtered step by step.

Finding data provenance and confirming the validity of data in LOD cloud after finding the desired data is another advantage of having such a metadata store. Since the first level datasets and their relationships in RDF triple are stored, this method enables users to identify trustworthy of the provenance information via these relationships. Regarding different definitions which exist in RDFS and ontology (such as owl:same as and so on) they may have different routes and ways from the data provenance and the found data. Since we can keep relations of datasets in the metadata store, we can also use best path algorithms to find effective way to reach desired data. To maintain relationship between datasets, many other methods including parent-child relationship can be used.

5 A Case Study

In a case study, we created a simple metadata store (an RDF file) about sample datasets such as Geography, Databases, Search engines, etc. Then, we used SPARQL for fetching data from the file and Java and Jena² framework as programming language in IntelliJ IDEA³ environment. When a user select a dataset group, we list all types of datasets that are related to it, and then, in the next step, the user can select the desired dataset among the list shown (Figure 2).

² <http://jena.sourceforge.net/>

³ <http://www.jetbrains.com/idea/>

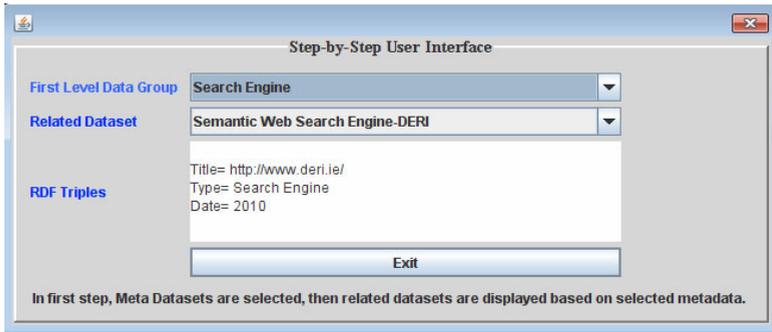


Fig. 2. A simple case study

6 Conclusion

In this paper, we considered a user interface of current Linked Data cloud and proposed a step-by-step user interface for easy and effective finding of data on LOD by designing a RDF metadata store for all datasets in LOD cloud. Having LOD metadata, users can be guided to find their data easily. On the other hand, when datasets metadata is created, valid data is delivered to users so that they can reach provenance information of datasets via the metadata. This is because it is possible to walk through datasets based on relationships defined in metadata store.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, Special Issue on Linked Data (2009) (in press)
2. Zembowicz, F., Opolon, D., Miles, S.: OpenChart: Charting Quantitative Properties in LOD. In: LDOW 2010 (2010)
3. Berners-Lee, et al.: Tabulator: Exploring and Analyzing Linked Data on the Semantic Web. In: Proceedings of the The 3rd International Semantic Web User Interaction Workshop (SWUI 2006) Workshop, Athens, Georgia (November 6, 2006)
4. Hastrup, T., Cyganiak, R., Bojars, U.: Browsing Linked Data with Fenfire. In: Hastrup, T., Cyganiak, R., Bojars, U. (eds.) Linked Data on the Web (LDOW 2008) Workshop, in Conjunction with WWW 2008 Conference (2008)
5. Margaritopoulos, T., Margaritopoulos, M., Mavridis, I., Manitsaris, A.: A Conceptual Framework for Metadata Quality Assessment. In: 2008 Proc. Int'l Conf. on Dublin Core and Metadata Applications, Berlin (2008)
6. Bose, R., Frew, J.: Composing Lineage Metadata with XML for Custom Satellite-Derived Data Products. In: SSDBM (2004)