RKBPlatform: Opening up Services in the Web of Data

Hugh Glaser
ECS, University of Southampton
hg@ecs.stoton.ac.uk

Ian C. Millard
ECS, University of Southampton
icm@ecs.soton.ac.uk

ABSTRACT

As the Linked Data initiatives and Web of Data become more widespread, sites that process and re-present the published data are growing in size and number. One challenge is to ensure that such sites do not themselves fall into the trap of failing to publish their new knowledge in a readily available manner. Not only should the work of such sites be re-published for Linked Data users, but it should also be accessible to site builders who have not yet embraced the Semantic Web. This paper presents the work that has been done with the RKBExplorer system to support this task, and describes examples of how it is used.

1. INTRODUCTION

RKBExplorer.com [1] is a Semantic Web application and infrastructure platform. The Explorer application provides a simple interface over multiple Linked Data sources to assist with the discovery and exploration of activities with the academic research domain, by providing search facilities, performing ontologically informed analyses to identify related resources, and assimilating digests of details for the resources being viewed.

To achieve this, the infrastructure provides a number of useful functions and tools. Information can be harvested from a number of sources, converted to RDF, and republished as Linked Data. The platform can access multiple Linked Data sources, both through SPARQL queries and fetching resources on demand from additional external data sets such as http://dbpedia.org/ and http://semanticweb.org/. Ontology mapping can be performed as necessary. Comprehensive facilities for identifying and managing co-references between entities are provided, in addition to methods employing "Community of Practice" style analyses to identify additional resources related to a given entity.

Clearly the information outputs and capabilities provided by the RKBExplorer system as a whole may also be of interest as individual resources to other Web of Data applications, or indeed ordinary web sites. We have recently been working to expose this information as services so that it can be used in this way.

The next part of this paper introduces some of the services, and the subsequent part describes two applications that utilise them.

2. RKBPlatform

Each of the components is exposed as a simple URI in a RESTful style. The type of result returned is constrained by the HTTP Accept header wherever possible, but can also be set by a parameter in the URI. As wide a variety of return formats as possible is offered, so that different types of consumers may easily access the services, including RDF in different representations and ontologies and non-RDF formats such as JSON. The following five services are the main ones currently being offered that are particularly useful to an application builder, although others are also available.

The current set of services and examples of their use can be found at http://www.rkbexplorer.com/demos/.

2.1 Co-Reference

A major part of the RKBExplorer system is that which deals with the knowledge that two URIs refer to the same non-information resource, managed by the CRS infrastructure [2]. The internal co-reference knowledge is available as an RKB service, in addition to the more public http://sameas.org/ service we offer that publishes data acquired using different (less conservative) policies. Typical URIs are http://www.rkbexplorer.com/sameAs/?uri=http://wiki.rkbexplorer.com/id/hugh_glaser, or an optional format can be specified: http://www.rkbexplorer.com/sameAs/?uri=http://sws.geonames.org/2647554/&format=rdf where format is *html*, *rdf* or *n3*. In addition the service accepts content negotiation of *text/html*, *application/rdf+xml*, *text/n3* and *text/plain*. Other parameters are possible, but are concerned with asking for visualisations or more data, so are not relevant to this paper.

2.2 Detail

This service gathers the RDF known about a particular resource from all the sources known via the co-reference service and formats it into HTML ready for display (using a Fresnel processor). The user or client can choose between a complete page (format=*page*), or an HTML fragment (format=*fragment*) if they wish to embed the information in another page. Typical URIs take the form http://www.rkbexplorer.com/detail/?uri=http://southampton.rkbexplorer.com/id/eprints-12614&format=page.

2.3 Network

The Network service provides a means of seeing which resources are connected to the resource under consideration:- if the resource is a person, then it can be considered to be a Social Network (concerned with professional connections) or a Community of Practice. The concept has been extended so that the Network of resources of other types can be seen, defined in a domain specific fashion. Results are ordered by how closely they are connected in a semantic sense, and are exactly as displayed in a normal RKBExplorer pane.

http://www.rkbexplorer.com/network/?uri=http://data.semanticweb.org/person/rudi-studer&type=person-researchArea&format=tsv is a typical URI.

The type can be any pairwise combination of *person*, *publication*, *mechanism*, *courseware*, *organisation*, or *researchArea*, and the format has a number of options: *tsv* (tab-separated variables) or *cloud* (HTML word cloud-style) and different RDF ontologies: *foafgroupxml* (FOAF groups in RDF+XML), *sioccommunityxml* SIOC community in RDF+XML) or *foafknowsn3* (FOAF knows in N3). Of course the RDF does not use the predicates correctly for different types of networks, but the service does not reject such requests at this time.

2.4 Explain

An important part of any system that purports to give knowledge is the ability to be told (in natural language) why the inference has been made. In the RKBExplorer this is achieved by clicking on the blue button with the question mark in it, and so this can also be used in other application, as can be seen in Figure 2. So the relationship between papers, for example can be found with the following URI: http://www.rkbexplorer.com/connections/?source=http://southampton.rkbexplorer.com/id/eprints-12614&target=http://southampton.rkbexplorer.com/id/eprints-

[15787&type=publication-publication](). The type parameter is the same as in the Network service above.

2.5   Search

This service offers the ability to make string searches across the *.rkbexplorer.com datasets, and any cached data from other sites.

The service needs a search string and the type of resource to search for, results are returned as a simple JSON structure.

Typical URIs are:
[http://www.rkbexplorer.com/explorer/search.php?type=person&search=Berners-Lee](), or perhaps for a publication, [http://www.rkbexplorer.com/explorer/search.php?type=publication&search=semantic+web]().

3.   APPLICATIONS
3.1   RKBGadgets

Google gadgets allow information providers to embed dynamic content into web pages or iGoogle home pages, or to place it on the desktop using Google Desktop or the OS X Dashboard.

Using the services described above, it should therefore be relatively straightforward for a third party to serve the data from the RKBPlatform in such Gadgets.

Mark Borkum, a postgraduate researcher in the School of Chemistry at the University of Southampton, has used the services exposed to develop a set of Gadgets for the RKB data, based around a single style focusing on people, publications, organizations or projects. Users can type in search strings and receive a word cloud of possible results; selecting one of these will cause the Gadget to display detailed information on the chosen resource (URI), with a word cloud of related resources of the same type below it. The word cloud also provides links to the Explain service, which indicates why resources are related.
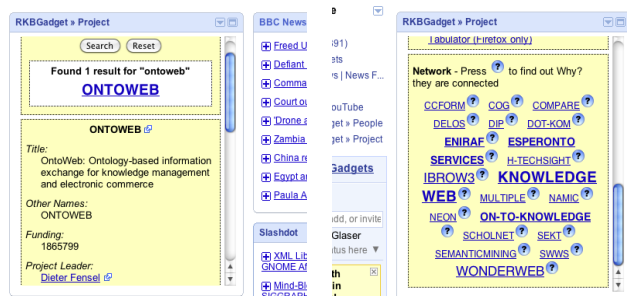

Figure 1. Two Clippings of the RKBGadget on an iGoogle page, looking at the Ontoweb Project

The RKBGadgets are freely accessible, and can be found online at [http://www.rkbexplorer.com/gadgets/]().

The source of the Gadgets is about 500 lines of Javascript, which includes all formatting and bookkeeping information, so it is fair to say that building the Gadgets on top of the RKBPlatform was indeed fairly straightforward. In fact as the exact nature of the required services becomes clearer as a result of this, it has become apparent that a more suitable set of services will make such application building even more straightforward.

3.2   Whozat

Whozat is an iPhone application that uses the RKBPlatform in a similar way to the Gadgets, albeit with an even simpler interaction in this prototype form.

The user can type in a person's name, and details of the most highly-connected person with that name will be displayed (similar to the Google "I feel lucky" button) as in Figure 2a. It is then possible to change the view to the view showing a word cloud of associated people (Figure 2b) or publications.

Figure 2. Two Screens of the Whozat iPhone App: (a) Search and Detail and (b) Word cloud of related people

The Whozat iPhone app is not sufficiently polished to be made available at present, but can be demonstrated. The code for the application is less than 100 lines of Objective C – clearly the application is not a finished product, but as with the Gadgets it demonstrates that appropriate services can be re-used in a very efficient manner.

4.   CONCLUDING REMARKS

The argument for Linked Data and the Web of Data is that data managers should publish their data in a machine-readable form. If this is true for the data providers, it must also hold true for the applications that process such data and build on it, especially when such application builders consider themselves Web of Data researchers. This paper has taken a number of steps to achieve this by offering RESTful services for the application components. There is more to be done:- offering the services in a more consistent fashion, and using and finding or defining ontologies in which to describe the results will help greatly. But we have shown that exposing the Web of Data services for re-use can be extremely powerful.

We challenge all builders of Semantic Web/Linked Data/Web of Data applications to expose their inferred data as Linked Data, in the same way that they in turn rely on the basic data providers to publish as Linked Data so that they can build their applications.

5.   REFERENCES

[1] Glaser, H., Millard. I.C., and Jaffri, A. 2008. RKBExplorer.com: A Knowledge Driven Infrastructure for Linked Data Providers. ESWC 2008, Tenerife, [http://eprints.ecs.soton.ac.uk/15152/]().

[2] Glaser, H., Jaffri, A., and Millard, I.C., 2009 Managing Co-reference on the Semantic Web, WWW2009 Workshop: LDOW2009, Madrid, [http://eprints.ecs.soton.ac.uk/17587/]().