

Linked Data for the masses: using open source infrastructure and the inbound/outbound Linked Data approach to bring added value to end user applications

George Anadiotis, Panos Andriopoulos, Dimitris Vekris, Aristotelis Zosakis
(IMC Technologies, Athens, Greece
ganadiotis@imc.com.gr, p.andriopoulos@gmail.com, dvekris@imc.com.gr,
zosakis@imc.com.gr)

Abstract: Linked Data is already gaining traction in the research community, however there is little to show to date in terms of practical Linked Data-powered applications and ready to use tools. We try to fill in this gap by taking a combined inbound/outbound Linked Data approach.

Keywords: Open source, Liferay, SILK, D2R, Tagging, Community, FOAF, SIOC, MOAT

1 Introduction

Aiming to explore the potential added value of Linked Data in building user applications, we can think of two ways that applications may interact with Linked Data: either as consumers or as providers. In the rest of the paper we shall refer to them as the inbound and outbound Linked Data approach respectively, show specific examples of both as well as how they can be combined. More details, as well as pointers to code for the Outbound Linked Data approach and to a sample dataset can be found at <http://62.38.158.77:2020/IMCTriplificationChallenge2009/>.

2 Inbound Linked Data: using external data sources to enhance applications

We chose to approach the inbound Linked Data scenario by capitalizing on existing datasets in order to solve a well-known issue in Web2.0 applications: tag ambiguity. Tags are free-form keywords used to annotate resources on the web; this means that the same tag can have different meanings in different contexts and it is hard to disambiguate between them. In order to solve this issue, we developed an application that grounds tag meaning by linking it to some existing dataset concept. The two ends of the spectrum in choosing which concept to anchor a tag to are using complex matching rules to automatically narrow down the selection, or trying to retrieve many potential matches and letting the user make the final choice.

We have experimented with different datasets (DBpedia, World Factbook etc), different ways of getting potential matches (direct SPARQL querying of available endpoints and using SILK¹) as well as different ways to boost performance (caching,

¹ <http://www4.wiwiw.fu-berlin.de/bizer/silk/>

indexing, query optimization). Our results so far have been very encouraging and we intend to test further with larger datasets and user groups.

3 Outbound Linked Data: mapping and making application and community data available

Our approach to outbound Linked Data is simply to provide the means for applications to make the data they produce available as Linked Data, thus enabling them to be reused and referenced elsewhere. Towards this goal we have chosen to create a Linked Data mapping for Liferay, a leading provider of open source enterprise portal and social collaboration software, as this is very widely used infrastructure with high impact. We used D2R to make information about online communities based on Liferay, as well as the content they produce, available as Linked Data. The goal was to map data stored internally in idiosyncratic formats to well-known vocabulary entities and use them as a common data exchange model to enable transparent querying and reuse of content across applications.

For this purpose, we performed a thorough mapping of Liferay's database to existing vocabularies such as FOAF, SIOC, Dublin Core, the Tagging ontology and MOAT. While the use of SIOC and FOAF was a natural match to describe user, community and content information and Dublin Core is generic enough to cover metadata aspects in virtually every domain, the Tagging ontology and MOAT is the point where the inbound and the outbound approach come together: tags are supported by Liferay, so we mapped them, gave them URIs and added the tag disambiguation information retrieved in the inbound approach to Liferay's database. Thus, we are able to make available data pertaining not only to tags and taggings but also tag meanings, using the MOAT vocabulary.

Since we expect this mapping to be used in production to publish large datasets, we paid particular attention to performance. Using D2R to publish large datasets has been shown to have some performance issues [Gray, 09], in particular when complex mappings that use joins are involved. For this purpose, in cases where data that makes up vocabulary entities had to be joined from different tables, we chose to do so using views and mapping entities directly to the views where possible.

In addition to making our module part of the official Liferay release, we also intend to develop additional modules to perform cross-site search and aggregation.

Acknowledgements

We would like to thank Bryan Cheung, Liferay Inc. CEO, for his support in our effort and his commitment to publish Liferay online community data as Linked Data.

References

[Gray, 09] Gray, A. J., Gray, N., and Ounis, I. 2009. Can RDB2RDF Tools Feasibly Expose Large Science Archives for Data Integration?. In Proceedings of the 6th European Semantic Web Conference on the Semantic Web: Research and Applications (Heraklion, Crete, Greece, May 31 - June 04, 2009).