RDFohloh, a RDF wrapper of Ohloh

Sergio Fernández

Fundación CTIC
Gijón, Asturias, Spain
sergio.fernandez@fundacionctic.org
http://www.fundacionctic.org/

Abstract. Data on the Semantic Web is modeled and represented in RDF. In the Social Web people usually do not give a further thought about this kind of formalism, whereas they do take care about the content. That is why it could be useful to have tools capable to export that amount of content in machine-readable formats, such as RDF. In this demo paper we present RDFohloh, a RDF wrapper of Ohloh, a Web 2.0 open source directory.

1 Introduction

The original vision of the Semantic Web [3], as a layer on top of the current Web, requires that data is published on the Web, and ideally linked with other useful resources. During the last years the Semantic Web community has made a big effort to make available more and more RDF datasets. Data can come from legacy sources, relational databases, or just making Web scraping; but also from social sources. Web 2.0 applications commonly provide some of its content via public APIs; so it is another big opportunity where data can be extracted and exposed as Linked Open Data [2].

2 Ohloh

Ohloh[1] is an open source directory. Its main goal is to aggregate projects and developers from any Web site. By retrieving data from revision control repositories (such as CVS, SVN, or Git), Ohloh provides statistics about the longevity of projects, their licenses (including license conflict information) and software metrics such as lines of source code and commit statistics. At this moment[2] Ohloh lists 15,532 projects and 23,430 developers. Another goal of Ohloh is providing a public RESTful API[3] with the most important information and many of that metrics.

[1] http://www.ohloh.net/
[2] Data retrieved on September 18th, 2008
[3] http://www.ohloh.net/api

3 RDFohloh, wrapping the wrapper

RDFohloh[4] comes to fulfill the requirement previously mentioned with Ohloh: consuming the data provided by its API and publishing it in RDF [11] as Linked Data. So since Ohloh could be considered a Web 2.0 wrapper for open source projects and developers, RDFohloh could be deemed a wrapper of the wrapper; the n-layers architecture in pure state.

3.1 Related work

Obviously the idea behind RDFohloh is nothing really new. There are several applications that export FOAF or SIOC[5] from other Web 2.0 applications. But for DOAP it could be summarized mainly in two:

- doap:store [12] is an online DOAP directory of computing projects, collaboratively built, where people do not need to register to the service, because it constantly retrieves decentralized projects description to build its database thanks to Ping The Semantic Web service.
- DOAPspace[6] is a registry/repository that contains DOAP scrapped data from several sources including SourceForge, Freshmeat and the Python Package Index.

3.2 Data and links

RDFohloh mainly uses three popular ontologies: SIOC [6] and FOAF [7] for users and DOAP [9] for projects. At the moment of this writing, RDFohloh is publishing 23,430 instances of `sioc:User`/`foaf:Person` and 15,532 of `doap:Project`. The result dataset has `skos:subject` links with DBpedia [1] concepts (for the moment only with the programming language of projects) and `owl:sameAs` links with DOAPspace projects.

3.3 Publication

One of the details specially attended in RDFohloh was how to publish the data. Using cool URIs [13] and content negotiation [10], it provides three views (RDF/XML, N3 and XHTML+RDFa) of each resource. All the data is published attending the best practice recipes [5], and the final result was successfully tested with Vapour [4].

[4] `http://rdfohloh.wikier.org/`
[5] `http://sioc-project.org/applications`
[6] `http://doapspace.org/`

4 Conclusions and Future Work

RDFohloh comes to expand the actual horizon of the Linked Data planet, providing social data from a rich source of information. However it is necessary to improve the project including some new features:

- Providing dumps of the all the data, properly described with Semantic Sitemaps [8], but first it is necessary to find how to cope with the limitations of the number of requests per day of Ohloh's API. With that dumps, it would be easier to provide also a SPARQL endpoint to query the dataset.
- Including source code metrics from Ohloh that now are missing in the RDF export, allowing possible semantic analysis of it.
- Improving actual links and add new ones to other open datasets.

All that features are in the roadmap of the project from its beginning, so hopefully it will be soon available.

References

1. S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In *Aberer et al. (Eds.): The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735, Busa, Korea, November 2007. Springer 2007.
2. T. Berners-Lee. Linked Data Design Issues. Available at `http://www.w3.org/DesignIssues/LinkedData.html`, 2006.
3. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 2001.
4. D. Berrueta, S. Fernández, and I. Frade. Cooking HTTP content negotiation with Vapour. In *Proceedings of 4th workshop on Scripting for the Semantic Web 2008 (SFSW2008). co-located with ESWC2008*, Tenerife, Spain, June 2008.
5. D. Berrueta and J. Phipps. Best Practice Recipes for Publishing RDF Vocabularies. Working Draft, W3C, 2008.
6. U. Bojars and J. G. Breslin. SIOC Core Ontology Specification. Member submission, W3C, 2007.
7. D. Brickley and L. Miller. FOAF Vocabulary Specification. Technical report, 2005.
8. R. Cyganiak, R. Delbru, and G. Tummarello. Semantic Web Crawling: A Sitemap Extension. Technical Report, DERI, 2007.
9. E. Dumbill. DOAP: Description of a Project. `http://usefulinc.com/doap/`.
10. K. Holtman and A. Mutz. Transparent Content negotiation in HTTP. RFC, IETF, 1998.
11. G. Klyne and J. J. Carroll. Resource Description Framework (RDF): Concepts and abstract syntax. Technical report, W3C Recommendation, 2004.
12. A. Passant. A user-friendly interface to browse and find DOAP project with doap:store. In *Proceedings of the 3rd workshop on Scripting for the Semantic Web (SFSW2007), co-located with ESWC2007*, Innsbruck, Austria, May 2007.
13. L. Sauermann and R. Cyganiak. Cool URIs for the Semantic Web. Interest Group Note, W3C, March 2007.