Practical Semantic Works – a Bridge from the Users' Web to the Semantic Web

Anca Paula Luca[1]

"Al. I. Cuza" University, Faculty of Computer Science,
General Berthelot, 16, Iaşi 700483, ROMANIA

Abstract. The PSW (Practical Semantic Works) script attends to enable collaboration between users and the semantic web by providing a method for users to take advantage of the semantic markup existent on the web with a natural language interface, while preserving the original semantic meaning upon usage of the retrieved information.

1 Vision

The problem with reaching towards the semantic web is that the semantic marked up data is not naturally accessible to users, it needs to be (heavily) preprocessed. Also, users are not used to, nor technically trained, nor willing sometimes to produce semantically marked up data, such that, without appropriate tools, the mixing between the users' (social) web and the semantic web is hurtful for one of the sides. The script proposed here attends to demonstrate how a bridge between the two can be built: a tool to search semantic knowledge bases, accepting queries in natural language and allowing users to reuse the responses while preserving the original semantic markup in the background.

2 Means

PSW is a script built for the Ubiquity [8] extension for Firefox, to approach the user in the browsing process. Ubiquity aims to provide a browser command line to access web services in natural language-like commands, allowing the developers to create new commands providing means to parse commands and to operate with the current page.

To provide responses to user's queries, semantically marked data from the Linked Data [4] project is used, accessed through SPARQL [6] and their respective SPARQL endpoints, so that the tool is able to provide answers from DBpedia [1], DBLP [2] or Linked Movie Database [3].

Finally, when a retrieved answer is to be inserted in the user edited document ((rich) text area), in order to preserve the semantic markup, since the user produced content is HTML (when not plaintext), we looked towards the microformats [5] initiative, which allows to add semantic markup to HTML data, using simple HTML markup conventions to represent specific types of data.

We modelled the Ubiquity search command as an RDF [7] triple with a missing component, its object. Such, for a command like `find <property> of`

`<subject> in <source>`, the `source` will designate the data repository to fetch answers from (for the moment DBpedia, DBLP Berlin and Linkedmdb), the `subject` search term will be used to look for a subject and the `property` will determine the predicate. The response will be a user accessible representation of the object(s) found for that subject and predicate. A flexible mapping mechanism is used to identify the subject and the property: each data source has a set of RDF properties to determine where the `subject` string is searched, and, upon success, the RDF property to be retrieved is determined from a mapping of the the user inserted `property`. Also, if a prior mapping does not exist for `property`, it will be searched as is, using a default per-source prefix. For the response, a predefined mapping between the retrieved property and a microformat is used to determine how to mark the response. In addition, a fourth parameter is accepted by the command, in the form `find <property> of <subject> in <source> as <format>`, which allows the user to specify the way the retrieved answer should be used (as a person, event, link, etc).

3 Further Thoughts and Conclusions

This prototype script can be extended to provide a mechanism to a allow users to define data sources and, by using an import mechanism or a discovery protocol, to make this accessible to non-technical users. Also, the usage of microformats can be improved by implementing an alignment mechanism between searched ontologies and microformats ontology, to determine automatically the format to be used for a retrieved property, as well actual fetch of the full data required by the particular format.

The PSW script aims to explore the space of applications which enable users to read and write the semantic web also (not only the 2.0 content web), in a transparent manner, without having to adapt to technical environment, and with no compromise from the data markup point of view: the "read" semantics are not lost upon "writing". Such a simple and fast approach, based on orchestrating existing data, tools and open standards, shows that, beyond the simplicity of the current stage, a full solution can be built in the direction of a bridge between the users' web and the data web.

References

1. DBpedia initiative: http://dbpedia.org/About
2. DBLP Berlin SPARQL endpoint: http://www4.wiwiss.fu-berlin.de/dblp/
3. Linked Movie Database: http://www.linkedmdb.org/
4. Linked Data initiative: http://linkeddata.org/
5. Microformats initiative: http://microformats.org
6. SPARQL Specification: http://www.w3.org/TR/rdf-sparql-query/
7. Resource Description Framework: http://www.w3.org/RDF/
8. Ubiquity Mozilla Labs project: http://labs.mozilla.com/projects/ubiquity/