Building the bases for a Semantic Web Browser

Davide Guidi, Laurian Gridinoc, Martin Dzbor and Enrico Motta

Knowledge Media Institute (KMi), The Open University, Milton Keynes, UK

Extended Abstract

In this paper we describe the first results of our efforts to build a solid framework for a Semantic Web browser, Power Magpie. With Semantic Web browser we envision an extension to a standard Web browser that augment its features with the ability to act on resources described in Web pages and to find resources semantically related to a Web page. Power Magpie differs from other proposed approach because it is not tailored to a predefined set of ontologies.

In our vision, the Power Magpie prototype will provide three main features:

1. The ability to automatically find and retrieve ontologies that are related to the generic Web page that the user is browsing.
2. A simple user interface to navigate these data (as well as the Semantic data available in the Web page itself, using technologies such as RDFa[3] or Microformats[1]).
3. The ability to provide services that exploit the discovered data.

The current state of the prototype already address the first two points, while the latter can be considered a further step that can be develop upon the proposed framework.

A high-level vision of the Power Magpie architecture is shown in Figure 1.

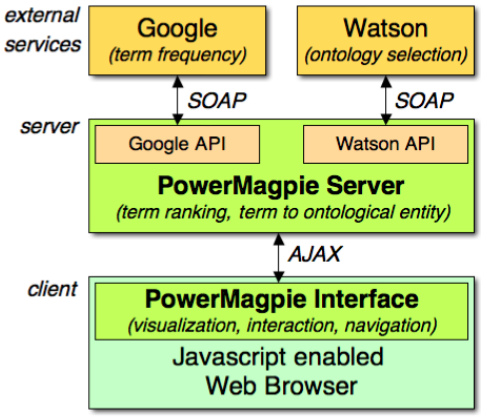

Fig. 1. The Magpie Architecture

The architecture consist of two main components: the *User Interface* and the *Power Magpie Server*. The User Interface is a Web browser extension that acts as an interface for the user. The server is instead the component that acts as a backend, providing the needed features. The two components can be run on the same computer or in different computers, in order to provide flexibility to the user.

In order to discover ontologies, Power Magpie can rely on external architectures such as Swoogle[5], SWSE[2], Watson[4] or ad-hoc peer to peer networks[1]. However, in the current prototype the main source of information is provided by the Watson architecture. Watson is a "gateway for the Semantic Web" that collects available semantic contents on the Web, analyzes it to extract useful metadata and indexes and implements efficient query facilities to access the data. From the point of view of Power Magpie, Watson is an ontology repository that provides access to the known Semantic Web.

We addressed several research problems due to the realtime requirements of the prototype. In fact, being a Web extension, the prototype must satisfy strict time constraints for the interaction with the user. Results obtained from a first series of experiments show that using the naive method of checking every term found in the Web page against the corpus of the ontology is obviously not feasible. We then applied several techniques in order to retrieve appropriate ontologies in due time:

- Exploiting the structure of the information found in Web pages (i.e. a word that appear in the title or in a header can potentially be more related to the context than a randomly chosen word in the text).
- Calculating weights for a set of words against the Google search results volume, in order to discover for example the *popularity* of these words in the "real world."

The result of our effort is a framework that retrieves ontologies related to a generic Web page in real-time. The current version of the prototype allows the user to explore the ontologies related to the Web page and to highlights concepts in the Web page. Two main tasks are already planned as future work. A first problem is to address the issue of the ambiguity that a can be generated by using the same word in different domains. The second main task is to provide services based upon this framework, exploiting the semantic content found in Web pages.

References

1. Microformats. Available online at `http://microformats.org`.
2. SWSE - Semantic Web Search Engine. Available online at `http://swse.org/index.php`.
3. B. Adida and M. Birbeck. RDFa Primer 1.0. Available online at `http://www.w3.org/TR/xhtml-rdfa-primer`, 2007.
4. M. d'Aquin, M. Sabou, M. Dzbor, C. Baldassarre, L. Gridinoc, S. Angeletou, and E. Motta. WATSON: A Gateway for the Semantic Web. 2007.
5. L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi, and J. Sachs. Swoogle: A Search and Metadata Engine for the Semantic Web. In *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*. ACM Press, November 2004.

[1] We are currently working on a peer to peer system for sharing ontologies using the software developed within the OpenKnowledge European Project (`http://www.openk.org`)

Explanation of the demo

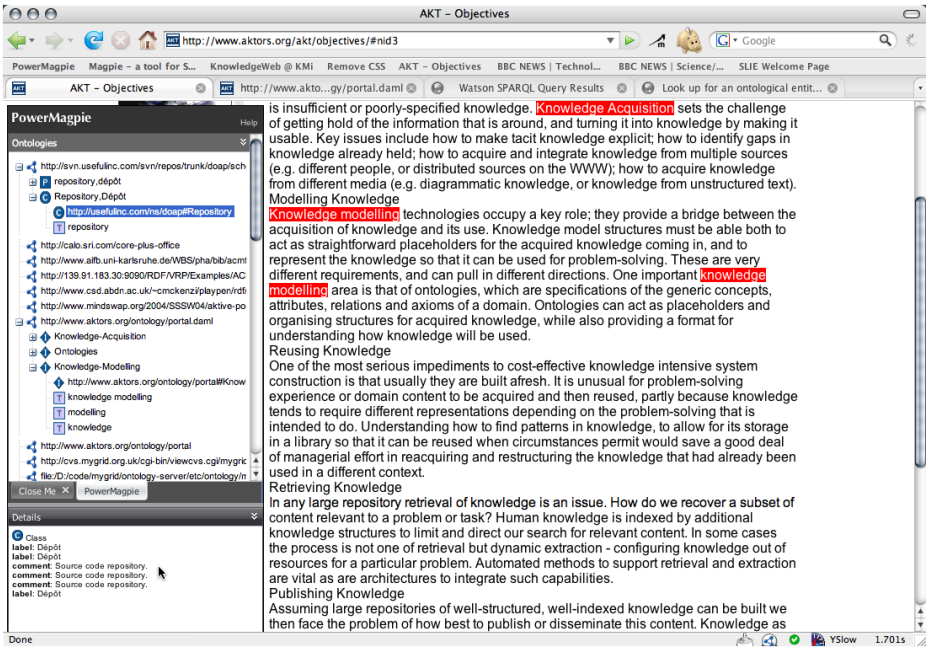

Fig. 2. A Power Magpie Screenshot

A screenshot of the Power Magpie prototype can be found in Figure 2. The figure shows the Power Magpie prototype run within the Firefox Web browser, triggered by a bookmarklet. In the picture the prototype already processed the current page's DOM, extracted and weighted terms against Google corpus, discovered, ranked and retrieved a set of relevant ontologies from Watson. These results are shown in the left sidebar, where the concepts matched in the Web page are grouped in the *Ontology view*. Other views enable the user to explore the data in different ways. The right sidebar contains the Web page selected by the user. If the Power Magpie is activated, then the words in the Web page that match concepts found in the ontologies are automatically highlighted. By clicking on the highlighted words, the user activates different views which contain more information about the retrieved ontologies and matched concepts and instances.