User Awareness in Semantic Portals

Ingo Brunkhorst, Nicola Henze

ISI - Semantic Web Group, University of Hannover &
L3S Research Center
Appelstr. 4, D-30167 Hannover, Germany
{brunkhorst,henze}@kbs.uni-hannover.de

Abstract. In this paper we discuss the benefits of *semantic* information portals. We argue that creating semantic information portals using Semantic Web technologies pays off right from the beginning: The possibility to reason about the (semantically annotated) Web resources allows for realizing personalization functionality which otherwise, i.e. compared to conventional portals, requires additional overhead. We provide a proof-of-concept for our claims within a semantic information portal for a research community and demonstrate personalization algorithms for improving user awareness.

keywords: semantic web, personalization, reasoning on the semantic web, semantic portal

1 Introduction

Information portals have been proven to be successful gateways to information in the World Wide Web: Information portals provide collections of relevant information on specific topics, group and structure the information, and support the user in retrieving, selecting and accessing electronic information. The idea of the semantic web as a layered architecture for realizing machine understandable meaning of Web resources holds particular new ways for building information portals: so called *Semantic Portals* can reason about the - now machine readable - semantic of Web resources, can search for relevant information in a more focused way by considering explicit semantic information, and can extract, rate and combine information resources in an advanced manner. In particular, the realization of user-adapted, *personalized* views on the data can profit from the provision of machine readable semantics, as user requirements can be more precisely considered in the retrieval, selection and presentation processes. At the current state, the creation of machine-readable semantic is not a fully automated process, but requires some overhead, e.g. in creating appropriate ontologies - for describing the application domain, for formalizing personalization attributes and requirements, etc. In this paper, we will demonstrate how the overhead of creating a semantic portal pays off right from the beginning. In Section 2 we outline the basic ideas of *semantic* portals, and demonstrate these ideas in the realization

of a semantic portal for the research community of the "REWERSE - Reasoning on the Web" project. In particular, we focus on the developed ontology which models the objects of discourse and the organization of research projects like partners, working goals and groups, co-ordination and participation, etc. This ontology is on the one hand the core model of the Semantic portal, and browsing the information in the portal is realized as projections of the information space according to (varying sets of) the concepts of the ontologies (see Section 2.2). In fact, the resulting portal for REWERSE could have been accomplished by using different techniques. But what distinguishes the approach of semantic portals from other portals realized by standard techniques is that we have advanced, nevertheless easy to realize ways to improve user support and to realize personalized views on the data. As an example, we demonstrate how user awareness can be supported in semantic portals. This awareness is based on a collaborative approach. Users browsing the semantic portal of REWERSE can easily grasp when other users are currently interested in similar information, or whether users with same interests are around. This is implemented by reasoning about the semantic descriptions of the Web resources with respect to the basic ontology of the portal, which models the community / project as a whole (see Section 3). We end the paper with a comparison of our work with related work on semantic portals, and a conclusion.

2 Semantic Portal for REWERSE

Semantic Portals are the result of a continued development of existing portal systems, basically *information portals* [15, 16] such as YAHOO[1] and DMOZ[2], providing access to an integrated and structured body of information about a specific domain. Additionally the *community-based* [15, 17] portals support the collaboration between members of a community, e.g. by allowing the contribution of information and news into the portal.

Crucial for every portal is the design of the navigation and search tools, because they provide the interface for accessing the information. Contrary to traditional portals using a hardwired navigation structure and text search, semantic portals exploit the properties and classification of information items and relationships between them, commonly using external ontologies [15]. Of course, the separation of structure and content of the portal is not new [11, 6, 3]. However, the Semantic Web approach allows for significant advancements: Ontologies describe objects of discourse, model the domain of interest, etc., thus allocate structure but also purpose of a semantic portal in machine readable format. Furthermore, this structure is interlinked with the Web, e.g. via the references of Web resources with respect to some ontology, but also via the connections between ontologies (see the efforts in ontology engineering, especially ontology mapping and ontology merging). This enhanced space of - for machines meaningful - information can be used for improved information syndication strategies and

[1] The Yahoo information portal at `http://www.yahoo.com/`
[2] The Open directory project at `http://www.dmoz.org/`

personalization strategies, realized in the logical layer of the Semantic Web architecture [2] by reasoning over these ontological information, evaluated against the personal information of users, retrieved from their user profiles.

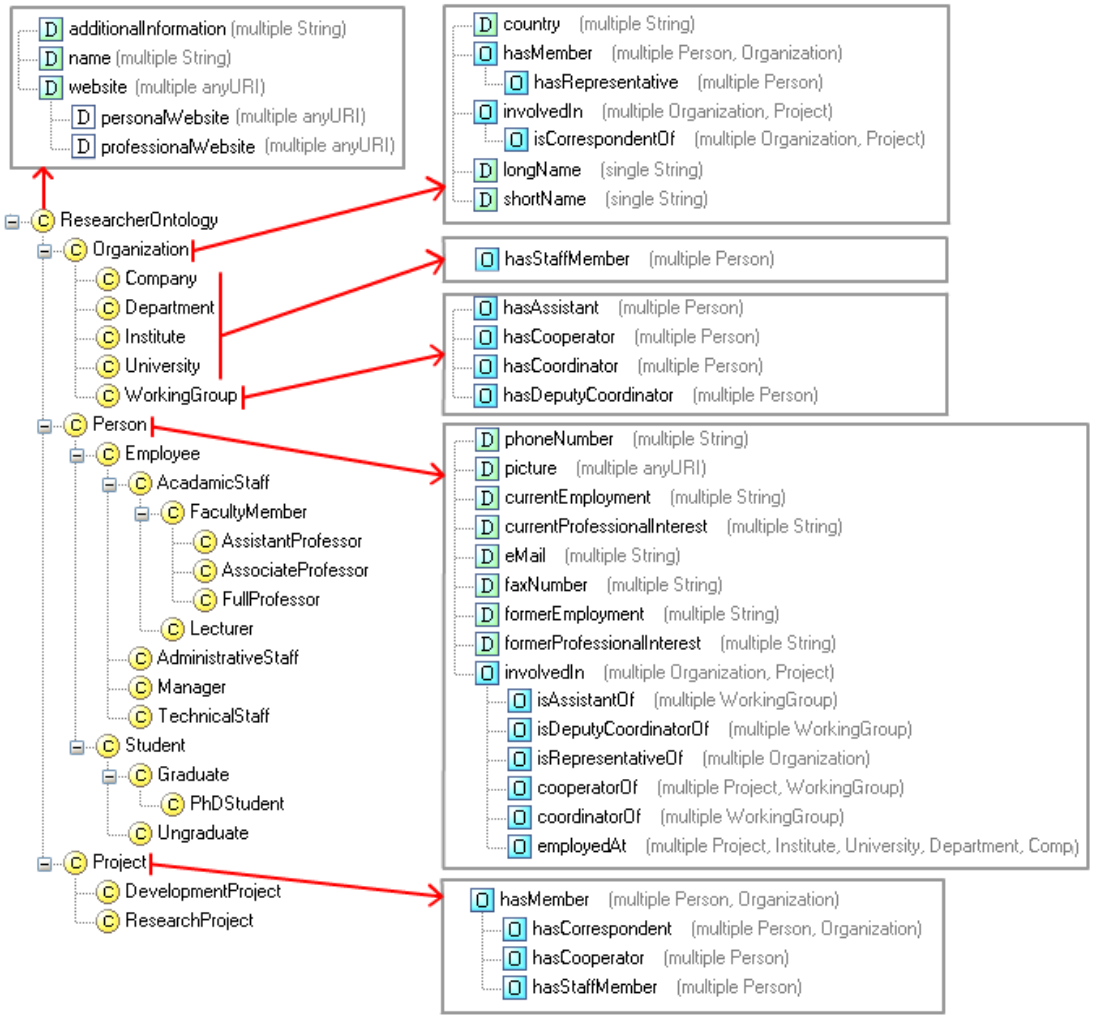

Fig. 1. Classes and Properties of the REWERSE-Ontology

2.1 Use-Case: A Semantic Portal for a Research Project

This section describes the development of a semantic portal for the European Network of Excellence REWERSE. We have constructed an ontology for de-

scribing researchers and their involvement in REWERSE. This "REWERSE-Ontology" has been built with aid of the Protégé-tool [14]. It extends the Semantic Web Research Community Ontology (SWRC) [18]. Like in the SWRC, the REWERSE-Ontology has three subclasses: *person*, *organization*, and *project*. Due to the extension of the SWRC, some more subclasses appear in it, e.g. university, department and institute as subclasses of organization. Additionally, the bibliographic metadata for publications is automatically retrieved from the members' public web pages, using the *Lixto* [1, 4] tool.

A synopsis of classes and properties used in the REWERSE-Ontology is depicted in figure 1. The current REWERSE-Ontology consists of 157 instances of persons and organizations using the 73 classes from the extended SWRC-Ontology. Figure 2 shows for an example how some classes and properties are instantiated.

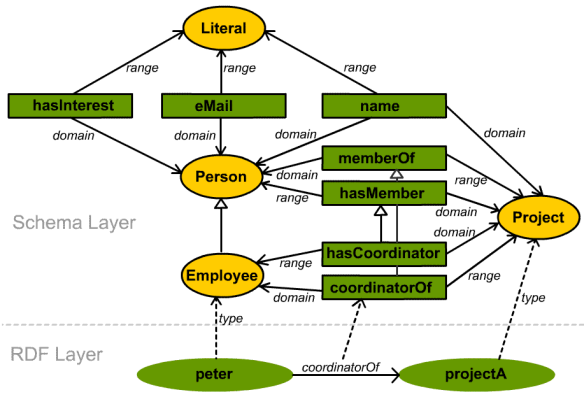

Fig. 2. Classes and Properties of the REWERSE-Ontology

2.2 Realizing the Portal

As the base of the REWERSE portal we use the SWED Portal technology. SWAD-Europe has been an European project run from 2002 to 2004. The Semantic Community Portal Group developed a prototype for a semantic portal, which we extended to fit our needs. As illustrated in figure 3, the portal itself consists of mainly two parts, the portal viewer and an aggregator component to import new data. The viewer application uses a Jena [9] RDF data model and technology from the Jakarta [8] project for generation of the interface. The aggregator scans known data sources for new or changed metadata and updates the data model accordingly. Rendering the content of the pages is based on templates and so called "facets", which are used to create browsable views of

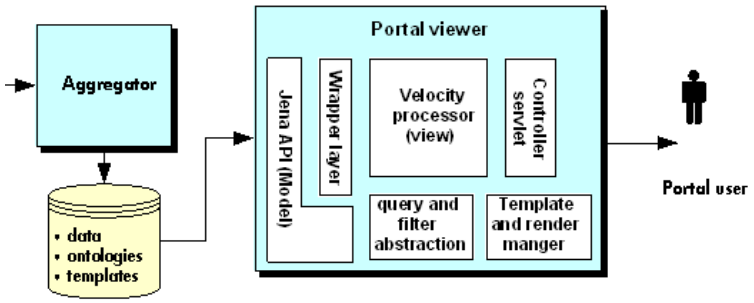

Fig. 3. Structure of the portal (Source: [7])

the underlying RDF data model, the REWERSE researcher ontology. Figure 4 shows the Portal starting page, including six facets, three of them based on the REWERSE ontology. The other three facets use a news data storage combined with the researcher ontology. In figure 5 the facet is used to select a certain view on the data, in this case it displays a list of persons, organizations and projects ordered by type. The template used to render the web page allows for further refinement of the selection, or browsing the ontology. For every resource in the RDF graph, a separate web page is generated, including all of the properties of the node. If a property is pointing to another resource of the ontology, the relation is visualized as an HTML link. The simple news system represented by the three other facets is the first extensions of the portal developed for the REWERSE project. This is complemented by a web based form to input new articles and to annotate them with metadata based on the project ontology, as well as an interface to the news aggregator module to manage the import of new articles.

3 Example: Awareness Module

The potential of having a semantic portal instead of generic text or HTML based content is the possibility to exploit the additional information with algorithms taking into account the relationships between the entities in the ontology. These algorithms have only to be developed once, and can be applied to other ontologies as well (Ontology Mapping [13]). Currently implemented are two example algorithms, exploiting a measure for "semantic distance" between the nodes in the graph of the ontology. The distance is computed by a component running in the portal system and exported as an XML document. A flash Applet, running on the client browser retrieves the document and display the distance as a Radar Screen animation, as shown in figure 6. However, for realizing these algorithms it is necessary to have some type of user authentication available, to map the user to the corresponding node in the researcher ontology and to

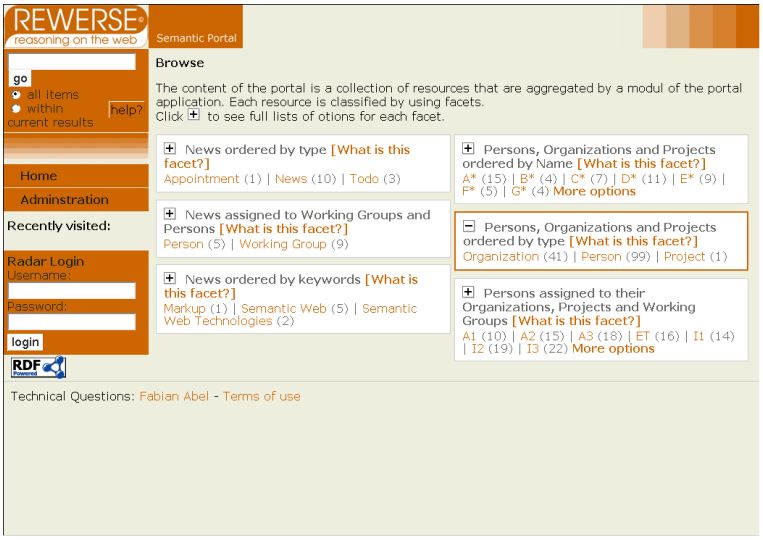

Fig. 4. Portal: Starting Page

identify the visited pages. The first algorithm computes the *browsing distance* of portal users. Since every node of the RDF graph is presented as a web page, this distance is measured by calculating the shortest path through the graph, between the pages currently viewed by the users of the portal. The algorithm for *professional distance* uses not the visited pages, but the nodes representing the authenticated users. Hereby the distance is calculated by comparing the affiliation with the projects working groups and organizations, e.g. persons working for the same work package are grouped closer together than those working in different groups.

4 Related Work: Semantic Portals

Already a few Semantic Portals and Tools have been developed and deployed on the Web. Tools used for the creation of ontologies and collecting metadata include Protégé [14] and Bibster [5]. Bibster is a Java-based system which assists researchers in managing, searching and sharing bibliographic metadata (e.g. from BibTeX files) in a peer-to-peer network. The bibliographic metadata for the REWERSE semantic portal is automatically retrieved from the web pages of the project partners using the Lixto tool [1][4], which originated at TU Wien, but is now distributed commercially by Vienna based Lixto Software AG. It is a system that assist the creation of rules and queries to extract metadata from web pages. For Semantic Portal Technologies, the REWERSE semantic portal will take advantage of the SWED portal software developed by the W3C and its partners in course of the SWAD-Europe Project [6]. Another initiative

Fig. 5. Portal Facet "Persons, Organizations and Projects ordered by type"

is SEAL [11], developed by AIFB from University of Karlsruhe. The research on semantic web and ontologies that was started with the European Project *OntoWeb* [12] is now continued and expanded in the project *KnowledgeWeb* [10], with a working group dedicated to designing and developing a portal for the project, in close cooperation with the REWERSE project.

5 Conclusion and Future Work

Realizing *semantic* information portals using recent semantic Web technologies brings new and fascinating possibilities for improving personalized access to Web resources. In this paper, we have discussed a demonstrator application: a semantic portal for the European Network of Excellence REWERSE. The core of this semantic portal is an ontology which models the world of research communities (the SWRC ontology), which we have extended to also model important aspects of research projects like working groups, the different roles of persons and institutions in such projects, etc. (the REWERSE ontology).

We have shown how personalization can be realized by using ontological information about Web resources: a simple but effective personalized user awareness support: the visitors of a portal are grouped according to their relation to the beholder: are they currently interested in similar topics (e.g. is the information they are currently viewing on similar topics), or do they normally have similar interests (e.g. are they working on similar topics as the beholder)? The beholder can easily check it via the implemented radar which visually depicts the distance of the visitors of the portal to her/him.

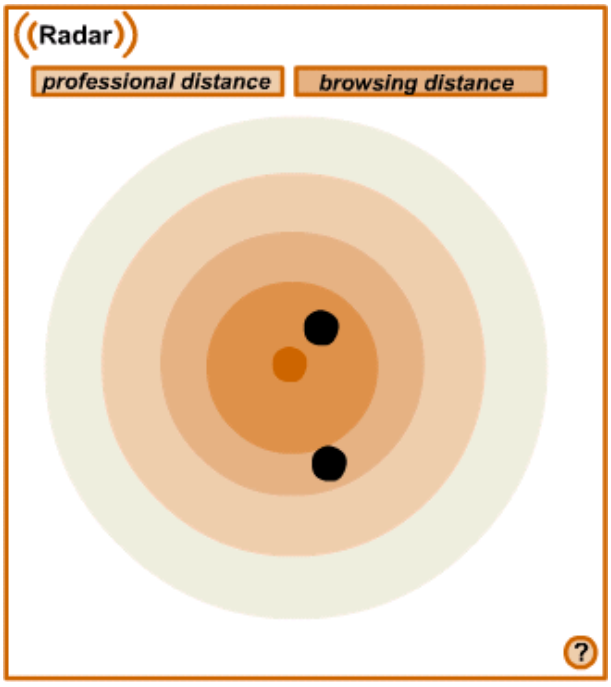

Fig. 6. Visualization of Distance

Acknowledgments

This work has been partially supported by the European Network of Excellence REWERSE - Reasoning on the Web with Rules and Semantics[3]. We would like to thank Fabian Abel for his work on the semantic portal for REWERSE.

References

1. Baumgartner, R., Flesca, S., and Gottlob, G. Declarative information extraction, web crawling, and recursive wrapping with lixto. In *6th International Conference on Logic Programming and Nonmonotonic Reasoning* (Vienna, Austria, 2001).
2. Berners-Lee, T. Semantic web - keynote at xml 2000 conference, 2000. http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html.
3. Domingue, J., and Motta, E. A knowledge-based news server supporting ontology-driven story enrichment and knowledge retrieval. In *Knowledge Acquisition, Modeling and Management* (1999), pp. 103–120.
4. Gottlob, G., Koch, C., Baumgartner, R., Herzog, M., and Flesca, S. The lixto data extraction project — back and forth between theorie and practice. In

[3] The REWERSE project at `http://rewerse.net/`

ACM Symposium on Principles of Database Systems (PODS) (June 2004), vol. 23, ACM.

5. Haase, P., Broekstra, J., Ehrig, M., Menken, M., Mika, P., Plechawski, M., Pyszlak, P., Schnizler, B., Siebes, R., Staab, S., and Tempich, C. Bibster — a semantics-based bibliographic peer-to-peer system.
6. Hewlett-Packard Development Company. W3c semantic web activity. ERCIM (for W3C), ILRT (University of Bristol), HP Labs, RAL/CCLRC, and Stilo, `http://www.w3.org/2001/sw/Europe/`.
7. Hewlett-Packard Development Company. Introduction to the swad-e portal structure, 2004. `http://www.swed.org.uk/swed/doc/portal-structure.html`.
8. The apache jakarta project, 2004. http://jakarta.apache.org/.
9. Jena - A Semantic Web Framework for Java, 2004. http://jena.sourceforge.net/.
10. Network of Excellence — Knowledge Web, 2004. http://knowledgeweb.semanticweb.org/index.html.
11. Maedche, A., Staab, S., Stojanovic, N., Studer, R., and Sure, Y. SEAL – A framework for developing SEmantic Web PortALs. *Lecture Notes in Computer Science 2097* (2001), 1–??
12. Ontoweb — ontology-based information exchange for knowledge management and electronic commerce, 2004. http://ontoweb.ontoware.org/.
13. OWL Web Ontology Language Guide, Feb. 2004. `http://www.w3.org/TR/owl-guide/`.
14. The protégé Ontology Editor and Knowledge Acquisition System, 2004. `http://protege.stanford.edu/`.
15. Reynolds, D., and Shabajee, P. Swad-europe deliverable 12.1.5: Semantic portals - requirements specification. Initial Release.
16. Reynolds, D., Shabajee, P., and Cayzer, S. Semantic information portals. In *Proceedings of the 13. WWW Conference* (New York, May 2004), vol. 13, ACM.
17. Staab, S., Angele, J., Decker, S., Erdmann, M., Hotho, A., Maedche, A., Schnurr, H.-P., Studer, R., and Sure, Y. Semantic community web portals. *Comput. Networks 33*, 1-6 (2000), 473–491.
18. SWRC - Semantic Web Research Community Ontology, 2001. http://ontobroker.semanticweb.org/ontos/swrc.html.