Realizing Semantic Web Portal Using Available Semantic Web Technologies and Tools

Lidia Rovan
Department of applied computing
University of Zagreb Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
lidia.rovan@fer.hr

ABSTRACT
Portals allow easy access to information by integrating heterogeneous applications or data sources in consistent way. It gives users a personalized and restricted view of domain information. Standard portal features could be improved employing semantic web technologies.
Although portals are now experiencing serious growth just as number of available semantic web tools, number of semantic web portals is negligible. In accordance to observed acceptance problems guidelines for developing semantic web portals are proposed.

Categories and Subject Descriptors
D.2.2 [**Software Engineering**]: Design tools and techniques – *object-oriented design methods;* H.1.2 [**Models and Principles**]: User/Machine Systems – h*uman factors;* I.7.2 [**Document and Text Processing**]: Document Preparation – *markup languages*

General Terms
Design, Human Factors, Standardization, Languages

Keywords
Semantic Web, Portal, Personalization, Single sign-on, Data integration, Data aggregation

1. INTRODUCTION
Due to the growing complexity of the Web, portals have become essential to the Internet users. Web portals provide users a single, integrated window or gateway to their personal world of information. The idea of a portal is to collect information from different sources and create a single point of access to data, expertise and applications. To provide users such functionality, portal deals with following technical problems: a) personalization, b) single sign-on, c) data aggregation (services), d) data integration, e) customization features.
The impact of semantic web technology to each of these problems is the issue of numerous scientific papers. Researches have resulted in various proposals for improving accustomed solutions. [1][2][3][5][6] Proposed solutions are ranging in width, and convergence to unique solution is still not perceived. Common solution is needed to finally attract developers to this area that would result in real practical solutions accessible to public.

We define a *semantic web portal* as any web portal that is developed based on semantic web technologies. We are in process of developing such web portal using available semantic technologies. Only standard technologies promising generic solution are selected. As a result we expect that we will be able to provide basic development guidelines in a form of portal architecture and design patterns. The reach of this kind of solution is questionable, so we will put accent on evaluating its usefulness and user satisfaction.

2. SEMANTIC WEB PORTAL POSSIBILITIES

2.1 Personalization and customization
Through sign-on mechanism users are identified and served with information customized for their needs and preferences. Performing personalization involves knowing the interests, activities, acquaintances or accessibility problems of a user. Besides handling explicit user preferences, personalization process implies monitoring user navigation, collecting requests and storing them in user profile, reasoning on them and on the domain ontology to deliver personalized content.
Semantically enriched personal profile imposes as an ideal solution. FOAF (Friend of a Friend) vocabulary is one of the most popular RDF (Resource Description Framework) vocabularies in the web and describes the basic information about persons and their relations that already made it particularly interesting for community portals.
There is lot of confusion in defining difference between customization and personalization. Customization is achieved by setting explicit view preferences by users while personalization is more based on user behavior. Semantic technologies strength is in identifying implicit rules, so obviously they do not bring a great advantage in the area of customization.

2.2 Single sign-on
Many portals today follow the principles of single sign on authentication mechanism that provides seamless access to different applications.
OpenID is an open, decentralized, free framework for user-centric digital identity; that is a decentralized mechanism for single sign on.
In the world of semantic web everything has a URI; hence every person has to have her own unique URI (identifier). In the latest FOAF recommendation, vocabulary has been updated to support usage of OpenID (a new property has been added - foaf:openid). FOAF used in combination with OpenID technology provides infrastructure that follows principles of the semantic web and supports: 1) authentication (OpenID protocol lets you prove that

you own a specific URL) and 2) user profile access in a shared and machine understandable way.
Other user profile management approaches are presented in [2].

2.3 Data aggregation and integration

There is a whole collection of data sources such as semi-structured textual documents (e.g. web pages) or structured data in the form of XML feeds, databases, web services and knowledge bases for almost any domain. Extraction of RDF data from XML and HTML documents is covered by GRDDL (Gleaning Resource Descriptions from Dialects of Languages) mechanism. GRDDL deals with Microformats and RDFa (Resource Description Framework attributes) encoding techniques. RDFa is probably a fastest excepted semantic web technology so far, it is more flexible and powerful than Micoroformats and it makes a perfect choice for semantic markup. XHTML 2 is a clean concise language that corrects many of web markup's past indiscretions. RDFa together with XHTML2 could bring a great benefit to web. As far as extracting data from relational databases into RDF is concerned, results are pretty clear. Among others, D2RQ [1]is a already widely used tool that serves its purpose.
Data aggregation problem boils down to problem of automated discovery, dynamic composition and execution of data fetching methods. Well-explored area of semantic web services deals with mentioned aspects. Research has resulted with W3C recommendation, usable APIs and tools for developing semantic web services, but practical examples are lacking behind. SPARQL endpoints need similar solution. In general, there is a need for repository that would contain data source descriptions made in ontologies for describing data sources. Until that occurs, we are forced to explicitly determine data sources.
Another difficulty is data integration that is reduced to problem of ontology matching.
The dominant use of public portals is still aggregating news feeds, so we will concentrate on supporting RDF/RSS aggregation and integration.

3. SEMANTIC WEB TECHNOLOGIES – LESSONS LEARNED

Motivation for this effort occurred during work on projects in the area of semantic web with students. It is observed that the students were significantly less productive in the tasks associated with the semantic web than in other areas. Even when working in this field was their choice, a decrease in their interest was significant. Majority of technologies is hard to learn and understand and lack of examples that would clarify things makes it even harder. It is necessary to develop simplified applications with the clear benefit and purpose to keep people interested in the area. Semantic web portal that we plan to develop should be clear, easy developed sample application.

4. STUDENT SEMANTIC WEB PORTAL

According to problems mentioned earlier, the idea is to build semantic web portal with technologies that are easy to use and that provide unambiguous immediate benefit. Student semantic web portal is place where students can access university's web-based service through single log-in, where they can communicate with other students and have personalized access to news and social events.
In order to satisfy today's users, web application must have a high level of interactivity, so it is an imperative to use Web 2.0 technologies. During our previous work some problems concerning implementation of semantic web technologies into highly dynamic application were discovered [4] so we expect to encounter integration problems. As we believe that creating and managing vocabularies is demanding area and should be left to specialists in that field, we have bounded only on usage of applicable existing vocabularies. Table 1 shows chosen technologies and their purpose within student semantic web portal solution.

Table 1 Chosen technologies

Chosen technologies	Purpose
FOAF	User profile, signing mechanism
OpenID	Signing mechanism
D2RQ	Mapping relational databases to RDF
vCard, iCal, geonames, lingvoj, DC, …	Domain ontologies
SIOC,	Forum
RDF/RSS	News feeds
RDFa	Mark up presentation data
Jena	RDF data manipulation

We expect that main contribution of our work will be in simplifying semantic web portal development process and providing a demonstrative example of current state in that area.

5. REFERENCES

[1] Acuña, C.J. et al.,Toward Web Portals Integration through Semantic Web Services, In International Conference on Next Generation Web Services Practices – NWeSP'05, 22-26 Aug, 2005, 6 pp.

[2] Breslin J.G., Grzonkowski S., Gzella A., Kruk S.R., Woroniecki T., Sharing Information Across Community Portals with FOAFRealm, Proceedings of the 4th IADIS International Conference on Web-Based Communities (WBC 2007), Salamanca, Spain, February 2007

[3] Henze N. , Abel F., User Awareness and Personalization in Semantic Portals, International Semantic Web Conference, 2005

[4] Rovan L., Nižetić I., Extending Java Web Applications for Semantic Web, in Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces, Cavtat, Croatia, 2008. 289-294

[5] Schmidt K, Stojanovic L, Stojanovic N and Thomas S, On Enriching Ajax with Semantics: The Web Personalization Use Case. ESWC 2007: 686-700

[6] Tziviskou C., Brambilla M, Semantic Personalization of Web Portal Contents. In the Proceedings of the 16th international conference on World Wide Web (WWW) 2007, Banff, Canada: p. 1245 - 1246.

[1] http://www4.wiwiss.fu-berlin.de/bizer/d2rq/