AN EVALUATION OF SEMANTIC WEB PORTALS

Rubén Lara, Sung-Kook Han, Holger Lausen, Michael Stollberg, Ying Ding, Dieter Fensel
Digital Enterprise Research Institute (DERI), Universität Innsbruck
http://deri.semanticweb.org/
Technikerstrasse, 13, 6020, Innsbruck (Austria)
{ruben.lara, sung-kook.han, holger.lausen, michael.stollberg, ying.ding, dieter.fensel}@uibk.ac.at

ABSTRACT

Web portals are entry points for information presentation and exchange over the Internet, used by a community of interest. Hence, they require an efficient support for communication and information sharing. Current Web technologies employed to build up these portals present serious limitations regarding information search, access, extraction, interpretation and processing. These limitations are naturally inherited by existing portals, thus hampering the communication and information sharing process between the community members. The application of Semantic Web technologies has the potential of overcoming these limitations and, therefore, these technologies can be used to evolve current Web portals into semantically enhanced Web portals. This paper presents the state of the art on the application of Semantic Web technologies to Web portals, points out the improvements achievable by the use of such technologies, and depicts requirements for future development of Semantic Web enabled Web portals. A wide coverage evaluation scheme and an evaluation criteria catalogue have been designed in order to consistently evaluate and compare existing Semantic Web portals.

KEYWORDS

Semantic Web, Web portal, Web community, information sharing, Web communication.

1. INTRODUCTION

The World Wide Web (WWW, or the Web for short), has made a huge amount of information electronically available, and is an impressive success story in terms of both available information and the growth rate of human users [Fensel and Musen 2001]. The Web has evolved from an in-house solution for around 1000 users in 1990 to more than 1 billion users and more than 1 billion documents (on the surface Web[1]), not only world-wide but also device-wide. This success has been based mainly on its simplicity, giving software developers, information providers and users easy access to new content.

So far, various communities have taken advantage of current Web functionality to strengthen communication and information exchange not only within the community but also with external communities or individual users. Various Web portals have appeared with the purpose of providing an open and effective communication forum for their members. In a prototypical case, a portal collects and presents relevant information for the community, and users can publish events or information to the community. Portals provide facilities for users to locate interesting information in the portal according to their personal preferences, topics, etc.

The same simplicity that made the impressive expansion of the Web possible has brought important, and in some cases critical, drawbacks that are hampering a further development of the Web. The general problem to find information on the Web is summarized in [Ding and Fensel 2002]: search results are imprecise, often yielding matches to many thousands of hits. Moreover, users face the task of reading the documents retrieved in order to extract the information desired. These limitations naturally appear in existing Web portals based

---

[1] If we consider pages generated dynamically and not found by traditional search engines (the deep Web), the number of available documents was estimated to be around 550 billion already three years ago [Bergman 2001]

on this technology, making information search, access, extraction, interpretation and processing a difficult and time-consuming task.

In this context, the Semantic Web [Berners-Lee et al. 2001] enables automated information access and use based on machine-processable semantics of data. Ontologies are the backbone technology for the Semantic Web and -more generally- for the management of formalized knowledge in the context of distributed systems. They provide machine-processable semantics of data and information sources that can be communicated between different agents (software and people). In other words, information is made understandable for the computer, thus assisting people to search, extract, interpret and process information.

Therefore, Semantic Web technologies can considerably improve the information sharing process by overcoming the problems of current Web portals. In this sense, portals based on Semantic Web technologies represent the next generation of Web portals.

We investigate the evolution of Web portals and analyze existing portals that make use of Semantic Web technologies. We restrict ourselves to Semantic Web portals (SW portal for short), which are defined as follows:

- It is a Web portal, that is, it is a Web site that collects information for a group of users that have common interests [Heflin 2003].
- It is a Web portal for a community to share and exchange information.
- It is a Web portal making use of Semantic Web technologies.

The aim of this paper is to compare existing SW portals regarding their features and underlying technologies in order to identify their strengths and weaknesses. Such a comparison requires a consistent evaluation criteria catalogue, which has been designed as part of our work and that is briefly described in the paper. The general purpose of our investigation is to show to what extent Semantic Web technologies have been applied to Web portals and which potential benefits these technologies have realized so far.

[Ding and Fensel 2002] conducted an extensive survey on existing ontology library systems. The coverage of this survey is very broad (including almost all the existing ontology library systems). However, its focus is limited to ontology management. [Maedche et al. 2001] proposed a generic approach for developing semantic portals, viz. SEAL (SEmantic portAL), that exploits semantics for providing and accessing information at a portal as well as constructing and maintaining the portal. Although the focus of this paper is different to our evaluation, this paper gave us a good starting point for designing our evaluation scheme. We further extend their proposed generic framework for SW portal development to include many other functions which we believe also relevant for a successful information sharing, such as functional ontology management (editing, browsing and searching, and versioning), Semantic Web Services [McIlraith et al. 2001], ontology-powered searching, and information processing workflow.

The paper is structured as follows: Section 2 presents the evaluation approach followed in the survey together with the criteria catalogue designed and used for our evaluation; Section 3 compares the main SW portals identified, using our criteria catalogue; Section 4 presents the conclusions of the evaluation and points out the future work.

2. EVALUATION APPROACH

The evaluation scheme used in our comparison is two-fold: first, it defines the structure a prototypical SW portal should present; second, it constitutes the base for our evaluation criteria. The evaluation criteria analyze the actual capabilities of a given portal compared to the required capabilities identified in our scheme. The definition of our evaluation scheme has been inspired by previous work on Semantic Web portals, especially by [Staab et al 2002].

In our approach, we distinguish three main scheme layers, summarized in Figure 1. First, the information access layer comprises the features for user-system interaction, which are consolidated in a usability evaluation of the portal user interface and an assessment of the portal as Web technology. Second, the information processing layer covers the way information items are processed by the portal. Five different steps are identified in this process, namely: creation, publication, organization, access and maintenance. How these steps are accomplished by a given portal is the focus of our evaluation for this layer. In addition, collaboration features are also included here in order to study what facilities the portal provides for user collaboration, e.g. expert finder. Third, the grounding layer encompasses the technologies supporting the features presented in the upper layers, split into Semantic Web technologies and system technologies.

The scheme layers adopt a functional perspective i.e. they describe the portal with regard to the functional features available for a portal user. The reason for this choice is that our main interest is to analyze to what extent the functionality of a traditional Web portal can be enriched by the use of Semantic Web technologies.

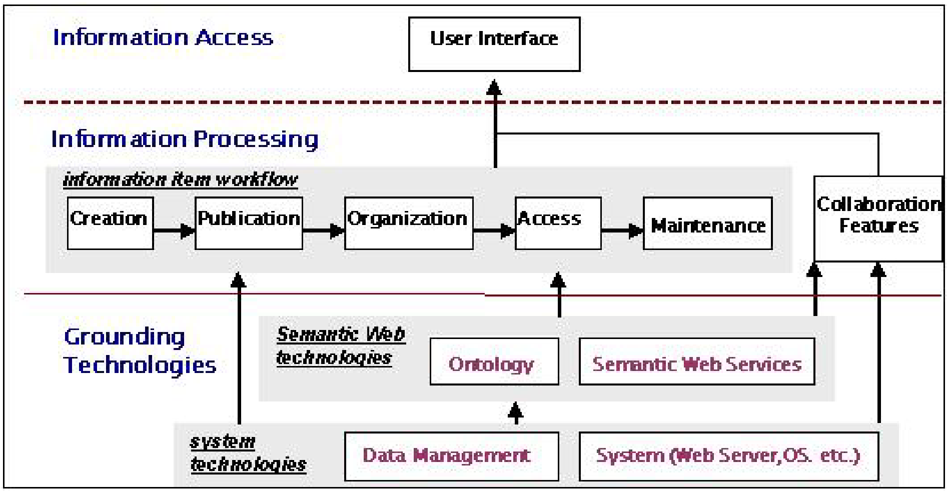

Figure 1. Evaluation scheme

A number of evaluation criteria have been identified for the three layers presented. They are summarized in table 1.

In the next section we use this criteria catalogue to compare existing SW portals. Although complete information has been gathered for all the portals analyzed, due to space reasons we do not present the details of each portal evaluation but a comparison of the selected portals regarding the main points of the three layers of our evaluation scheme.

3. PORTALS EVALUATION

In our comparison, we have identified a number of Web portals that apply Semantic Web technologies to enhance their information sharing capabilities. A detailed evaluation has been accomplished for the SW portals that a) employ these technologies and b) fit to our definition of SW portal. We have inspected two academic portals (Esperonto[2] and OntoWeb[3] portals) and two commercial portal technology infrastructures (Empolis K42[4] and Mondeca ITM[5]). In this section, we will provide a comparison between these portals following the three layers of our evaluation scheme. Other portals have been identified, such as the SWWS portal[6], the Mindswap portal[7], KA2[8], parts of the AIFB portal[9], the KAON portal[10] and the OntoWeb Edu

[2] http://www.esperonto.net/
[3] http://www.ontoweb.org/
[4] http://k42.empolis.co.uk/
[5] http://www.mondeca.com/english/
[6] http://swws.semanticweb.org/
[7] http://owl.mindswap.org/
[8] http://ka2portal.aifb.uni-karlsruhe.de/
[9] http://www.aifb.uni-karlsruhe.de/Personen/
[10] http://kaon.semanticweb.org/

portal[11]. Nevertheless, these portals only make a partial use of Semantic Web technologies and, therefore, they are not included in the comparison.

Table 1. Evaluation criteria catalogue

<table>
<tr><td></td><td colspan="3">Evaluation criteria</td><td>Description</td></tr>
<tr><td rowspan="18">Grounding technologies</td><td rowspan="5">SYSTEM TECHNOLOGIES</td><td rowspan="3">Data management</td><td>Data storage</td><td>Data storage devices and type of information stored in them.</td></tr>
<tr><td>Sorting and indexing</td><td>Sorting and indexing techniques used to improve data storing and retrieving capabilities.</td></tr>
<tr><td>Data transfer</td><td>Data formats and transmission protocols used in the system.</td></tr>
<tr><td rowspan="2">System maintenance</td><td>System administration</td><td>Portal Administration options.</td></tr>
<tr><td>Security technology</td><td>Technologies to ensure secure access to the information in the portal.</td></tr>
<tr><td rowspan="13">SEMANTIC WEB TECHNOLOGIES</td><td rowspan="3">Ontologies</td><td>Ontology type</td><td>Domain and application ontologies used in the portal, and their purpose.</td></tr>
<tr><td>Ontology structure</td><td>Ontology structure and size.</td></tr>
<tr><td>Additional facets</td><td>Internationalization, multilingualism, balance of expressivity and scalability of the used ontologies.</td></tr>
<tr><td colspan="2">Inference and reasoning</td><td>Inference and reasoning mechanisms used in the portal.</td></tr>
<tr><td rowspan="4">Ontology management</td><td>Edition</td><td>Edition facilities for the ontologies used.</td></tr>
<tr><td>Maintenance and versioning</td><td>Ontology maintenance and versioning capabilities.</td></tr>
<tr><td>Ontology search for administration</td><td>Support for finding specific ontologies or part of the ontologies used in the portal, for administration purposes.</td></tr>
<tr><td>Standardization</td><td>Semantic Web ontology languages used.</td></tr>
<tr><td rowspan="2">Semantic Web Services</td><td>Functionality</td><td>Functionality of the portal accessible via Web Services.</td></tr>
<tr><td>Automation support</td><td>Web Services automation support.</td></tr>
<tr><td rowspan="6">Information processing</td><td colspan="3">CREATION</td><td>Creation of new information items and assignment to the appropriate ontologies.</td></tr>
<tr><td colspan="3">PUBLICATION</td><td>Publication of information items to make them accessible to the community.</td></tr>
<tr><td colspan="3">ORGANIZATION</td><td>Storing and indexing capabilities of the information items.</td></tr>
<tr><td colspan="3">ACCESS</td><td>Retrieval functionality, mainly search capabilities.</td></tr>
<tr><td colspan="3">MAINTENANCE</td><td>Maintenance of created information items.</td></tr>
<tr><td colspan="3">COLLABORATION FEATURES</td><td>Additional features to support information sharing and communication.</td></tr>
</table>

[11] http://qmir.dcs.qmul.ac.uk/ontoweb/index2.html

Information access	**USABILITY**		Usability of the portal.
	GENERAL ASSESSMENT AS WEB TECHNOLOGY	**Coverage**	Relevance and completeness of the information in the portal.
		Maturity of implementation	Maturity of implementation of the current version of the portal.
		Personalization	Personalization features to improve information access.
		Reliability of information resources	Completeness and consistency of the information published in the portal.
		Help and documentation	Help and documentation available for the portal users.

3.1 Grounding technologies

In the grounding technologies layer, we evaluate the system technologies and the semantic technologies used by the different portals to support the upper layers. The next sections give a comparison between the evaluated portals regarding these technologies.

3.1.1 System technologies

Most of the portals use the traditional three tier architecture: a database and / or a file system as a backend data storage layer, a Java Servlet based user interface for the front-end, and various server components in the middle tier. Regarding document storage, only OntoWeb leverage an existing document management framework functionality, building on top of ZOPE[12]. Other evaluated portals just provide simple upload functionality and use the Web servers' file system (if at all).

Data transfer has been achieved by either using existing protocols (such as SOAP) or home-made solutions like passing serialized Java Objects directly over TCP/IP. None of them has used a fully Service Orientated Architecture (SOA), not employing (Semantic) Web Service technology for the communication between components (internal and external).

Systems are administrated directly via various application servers and operating system mechanisms. Security in the communication of the information is mostly achieved by providing password-protection for private areas, using methods offered by the application server.

3.1.2 Semantic Web technologies

Semantic features are currently implemented in a limited way, for example only providing taxonomy import and export features. The reasons for this limited exploitation of Semantic Web technologies could be their immaturity and the difficulty of employing them due to technical reasons.

The ontologies used in the evaluated portals are normally specifically developed for the portal, even though some of them are reusing existing ontologies e.g. the OntoWeb ontology relies on the KA2 ontology. The ontologies' character is rather static than dynamic and updates are only allowed to a limited extent; updates simply overwrite existing ontologies and very limited versioning mechanisms are used. Only the Mondeca ITM technology offers multi-language support for its ontologies. Inference or reasoning is very limited, mostly restricted to simple inverse, transitive or symmetric properties of ontological concepts or relations. Only the Esperonto portal makes use of different ontologies, while the other portals evaluated use a single ontology.

The control of ontology data (information items) is usually handled by defining different user levels. Normally these levels are: portal administrator (full rights), registered portal members (some rights), and guest visitors (limited rights). Ontologies and instances are maintained separately by using existing ontology editors such as WebODE (Esperonto portal), or by using home-made solutions such as OIModeller (OntoWeb), WebAuthor and Ontogen (Empolis K42) and the ITM editor (Mondeca ITM). Empolis K42 and Mondeca ITM are restricted to home-made editors after importing the first version of an ontology. Most portals (Esperonto, OntoWeb, Mondeca ITM) support multiple formats for the initial ontology creation and

[12] http://www.zope.org/

for exporting of schema and instance data. Some heuristic rules are added to achieve consistency e.g. when a concept is deleted from the ontology, its instances become instances of its super class. None of them provide a sufficient versioning mechanism to trace changes between different versions of the ontologies.

For the internal representation of the ontologies, the academic portals mainly use RDF. The case of the commercial products is different, as they are mainly based on the Topic Map [Biezunski et al. 1999] paradigm. It is worth to notice that Mondeca ITM uses OWL for ontology description, uses RDF Dublin Core for document annotation, and Topic Maps/RDF for the interaction with the client. The expressiveness of Topic Maps is usually equivalent to a taxonomical structure with relations. More advanced ontological modeling features like cardinality constraints and inference enabling properties, e.g. inverseOf are only partly included (OntoWeb).

Although Semantic Web Services are one of the most relevant applications and exploitation areas for Semantic Web technologies, none of the portals evaluated implements or supports them. Future plans to implement (Semantic) Web Services are mentioned for every portal, but no concrete strategies and technologies to be used are detailed.

Technically, the challenge is to reuse work from both the ontology and the document management areas. Therefore, the harmonization and integration of Semantic Web technologies with existing Content and Document Management Systems and, furthermore, making them an inseparable unit is one the main tasks to be accomplished.

3.2 Information processing

The information processing facilities of a SW portal consist of five life cycle stages: creation, publication, organization, access and maintenance. Each portal varies in the implementation of these phases and the borders are often fuzzy. However, classifying the processing into these steps provides a better base for our comparison.

In all the portals investigated in detail, the creation of a new information item is based on HTML forms that present the attributes of the corresponding ontology concept. K42 is the only exception, as it does not provide complete Web based forms for end users, which makes the creation of new items by end users more difficult. The Esperonto portal provides Web-based forms to guide the creation process, but only the item identifier and the item description are given at this step. Further item information is given by editing the instance after its creation. We have found this two-step process rather contra intuitive, and it introduces an unnecessary second step increasing the creation time. OntoWeb does provide complete forms and in addition it integrates pre-defined ontologies such as DC Core and BibTex, but only ITM also enhances this step with automatic features such as extracting the author name directly from the meta-data of a Microsoft Word document. In every portal, the assignment of a new item to the ontology is done implicitly when a new item is created. The item is assigned to the ontology concept selected by the user at the beginning of the creation step. For the creation of documents, every portal relies on external editors like Microsoft Word.

The publication of a new information item is usually split into two steps: the submission by the creator and the validation by a portal administrator. The only exception is the Esperonto portal, which publishes the item right after its creation, not including any validation step. The actual sub-steps in the publication phase vary between the portals, mainly depending on the number of different user levels defined. For example, K42 permits publication only by administrators. Within our evaluation we found no evidence that any portal interweaves Semantic Web technology into its publication process, not including, for example, publication rights over ontology concepts.

For the access of the information by the portal users, most of the portals provide ontology based navigation or browsing functions. ITM does combine this with a full text search of the information item content. OntoWeb does not offer full text search and does not interweave the ontology with a thesaurus. K42 offers a wide range of visualization tools, but does not provide an ontological search form as defined in our criteria. Esperonto includes ontological search, but it includes no graphical representation and we found several errors when combining ontological search with keyword-based search.

Regarding the organization of information, OntoWeb maps its ontology to a simple object model (ZOPE objects), which does not support rich features like other ontology repositories. Both K42 and ITM use self developed repositories, and only the Esperonto portal employs a pre-existing repository with rich functionality (WebODE).

Maintenance is closely related to the organization of information items. Support for collaborative evolution of the ontology schema and easy modification of instance data should be provided by a mature SW portal. Nevertheless, none of the portals have elaborated on evolution concepts for the schema. The underlying ontology system of the Esperonto portal does support versioning, but this feature is currently not exploited within the portal. Furthermore, the Esperonto portal has serious limitations regarding the maintenance of instance data: it does not allow any user to delete published items, and the portal administration has to be contacted when an item needs to be deleted. OntoWeb does not support versioning but has a good integration of changes into the publication workflow. K42 and ITM just overwrite existing information when modifying instance data.

3.3 Information access

Regarding the purpose of the portals evaluated, we found that the academic portals are used as the document management and dissemination point for research projects, while the commercial portals are intended to be instantiated by customer developers for their application to different areas, e.g. a conference portal or a portal solution for Knowledge Management (see demonstration sites of K42 and Mondeca ITM). The portals' use is mainly limited to the creation, maintenance and access of information about their application domains. Functions to facilitate user's communication are ignored, such as discussion forums, mailing list archives or referential materials.

The content provided by the portals inspected only covers its application domain, giving no attention to related areas. Regarding the maturity of implementation, we found a general lack of maturity of the implementations evaluated. With the exception of OntoWeb, the portals do not allow personalization of information spaces. Basic help and documentation are provided to facilitate an easy use of the portal, with the exception of the Esperonto portal, which does not offer any documentation or help for the portal users. Nevertheless, we found the documentation and help provided by the other portals required improvement, especially for guest visitors.

Regarding navigation, the evaluated portals rely on their conceptual models, i.e. the underlying ontologies. E.g. they render tree structures according to the conceptual model behind.

4. CONCLUSIONS

One of the main benefits of the Semantic Web portal approach is the ability to model a portal structure using ontologies as the starting point. Ontologies are best suited to represent consensus knowledge and its structure. This is exactly what is needed to exchange information within a community of interest and to enable automated processing of information items. Therefore, the use of ontologies provides a solid starting point for the construction of community portals. Conventional portals try to solve this with various structuring methods like content type, view, etc. but they often lead to user confusion and compatibility problems with other portals. Furthermore, these conventional information structures do not enable the automatic generation of portals from the knowledge model. An essential benefit of SW portals is that they are able to load an initial ontology and build a system out of the box that can satisfy the user needs.

Nevertheless, in our evaluation we found three serious limitations of current SW portals. First, existing semantic portals do not employ multiple ontologies and they are not designed to interoperate with related portals, not paying any attention to ontology alignment techniques to enable information sharing. This is a serious shortfall that must be improved in order to realize the promise of the Semantic Web, that is, interoperation. Semantic Web portals should evolve current monolithic Web portals into P2P portals, sharing information and improving their information export and import capabilities. SW portals must be able to deal with two different (but related) sources of heterogeneity: heterogeneity in terms of multiple ontologies and heterogeneity in terms of multiple portals. Only the OntoWeb portal interacts with related portals (OntoWeb Edu and OntoWeb Roadmap). Nevertheless, this interoperation is achieved using an ad-hoc procedure, and the OntoWeb portal was initially designed without considering interoperation issues.

Second, we found a big lack of versioning capabilities. In a dynamic environment such as the Web, and in dynamic communities, the ontologies underlying the evaluated portals may change over time. For that reason, it is essential that SW portals make use of versioning mechanisms in order to keep the information consistent and to make changes traceable. In the evaluated portals only very limited versioning capabilities are used.

Third, the existence of strong community features is essential for the information sharing process. The analyzed portals do not present sufficient functionality to support an appropriate communication between the portal users.

In addition to these major shortfalls, the following features must be considered in order to bring SW portals to their full potential:

- Existing portal and content management technologies should be reused, not reinventing the wheel for problems already solved, e.g. user management or publication workflow.
- Multi-functional search: different search levels should be provided by a SW portal, such as keyword-based search (including search over documents), visual ontology browsing, ontology powered search and inference-powered search. These different levels of search provide great search flexibility and exploit appropriately SW technologies to improve search results. Existing portals only provide partial multi-functional search capabilities.
- Standardization: SW portals should adopt current standards as much as possible to facilitate the communication and re-usage of knowledge structures i.e. ontology schema data, and information items i.e. ontology instances.
- Future development of successful SW portals should also focus on Semantic Web Services. Semantic Web Services can expose the portal functionality to be accessed programmatically, and this can constitute a basic point to achieve interoperability between portals.

To develop full-fledged SW portals that make use of all the benefits of Semantic Web technologies, the limitations encountered in our evaluation must be solved. In our future work, we will track the evolution of the evaluated portals and identify the new SW portals that may appear.

ACKNOWLEDGEMENT

This work has been partially funded by the EU projects OntoWeb (http://www.ontoweb.org/), SWWS (http://swws.semanticweb.org/), COG (http://www.cogproject.org/) and h-TechSight (http://prise-serv.cpe.surrey.ac.uk/techsight/). We would like to thank the members of DERI Innsbruck for their comments and support, and especially to Ann Johnston (DERI Galway) for proof reading the paper.

REFERENCES

Bergman, M. 2001. The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing.* Volume 7, Issue 1.

Berners-Lee, T. et al, 2001. The Semantic Web. *In Scientific American*, May 2001.

Biezunski, M. et al, 1999. Topic Maps: Information Technology - Document Description and Markup Languages. *ISO/IEC 13250:2000.*

Ding, Y. and Fensel, D, 2002. Ontology Library Systems. The key to successful Ontology Reuse. *Proceedings of the First Semantic Web Working Symposium.* California, USA, pp. 93-112.

Heflin, J, 2003. Web Ontology Language (OWL). Use Cases and Requirements. *W3C Working Draft,* 31 March 2003.

Fensel, D. and Musen, M.A, 2001. The Semantic Web: A Brain for Humankind. *In IEEE Intelligent Systems,* March/April 2001, pp 24-25.

Maedche, A. et al, 2001. SEAL – A Framework for Developing Semantic Portals. *Proceedings of the 18th British National Conference on Databases*. Oxford, UK.

McIlraith, S.A. et al, 2001. Semantic Web Services. *In IEEE Intelligent Systems.* March/April 2001.

Staab, S. et al, 2002. AI for the Web – Ontology-based Community Web portals.