Semantic web portals: state-of-the-art survey

Holger Lausen, Ying Ding, Michael Stollberg, Dieter Fensel, Rubén Lara Hernández and Sung-Kook Han

Abstract

Purpose – To present the state of the arts application of semantic web technologies in web portals and corresponding achievable improvements for identifying the potential improvement made by semantic web technology.

Design/methodology/approach – An evaluation scheme is proposed to investigate various web portals that make use of semantic web technologies in order to identify their strengths and weaknesses. This scheme consists of three layers: information access, information processing and grounding technologies. Two academic portals and two commercial portals are selected based on the definition of semantic web portal. Detailed evaluation based on the proposed scheme is conducted on these four select portals.

Findings – Semantic web technologies can definitely increase the information consistency and the information processing quality of web portals by using ontologies to model portal structure and consensus knowledge. Furthermore, semantic web services will be acting as the key technologies to lift current portals to next level.

Originality/value – This paper proposes an elaborated evaluation method for investigating various portals. It reveals the current status of semantic web applications in web portals.

Keywords Semantics, Portals, Computer applications, General dissemination of information, Communication technologies

Paper type Technical paper

1. Introduction

Various communities have taken advantage of the current Web functionalities to strengthen communication and information exchange inside and outside of the community. Miscellaneous web portals have appeared with the purpose of providing an open and effective communication forum for their members. Semantic web (Berners-Lee et al., 2001), as the next generation web, 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. Semantic web technologies can considerably improve the information sharing by overcoming the problems of current web portals. In this sense, portals based on semantic web technologies represent the next generation of web portals.

In this paper we investigate the state of the evolution of web portals and survey existing portals that make use of semantic web technologies. The scope of portals investigated is restricted to semantic web portals (SW portal for short), which are defined as follows:
- It is a web portal. A web portal 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 based on 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. A general purpose of our investigation is to show to what extent semantic web technologies are applied to portals at this point of time and which potential benefits these technologies have been realized so far.

The paper is structured as follows: Section 2 presents the evaluation approach followed in the survey; Section 3 analyzes the survey results and provides a comparison among these portals; Section 4 indicates related work and Section 5 concludes the survey and points out the future work.

2. Evaluation scheme

Figure 1 shows the scheme that is utilized for describing and evaluating SW portals in this paper. It mainly distinguishes three layers: information access from the user's perspective, information processing features of the portal and the grounding technologies.

**Grounding technologies layer**

It contains the basic technical building blocks of a SW portal. These are system technologies and semantic web technologies used in portals:

**System technologies.** For evaluation of an SW portal, information about used technologies is provided, followed by a more detailed description on data management and system maintenance techniques. Through such description, an overall functional understanding of the system shall be attained:

---

**Figure 1** Semantic web portal layers

<table>
<thead>
<tr>
<th>Information Access</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Usability</strong></td>
</tr>
<tr>
<td>Assessment of Web Technology</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>collaboration features (synchronous, asynchronous)</td>
</tr>
<tr>
<td>Information Item workflow</td>
</tr>
<tr>
<td>Creation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grounding Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic Web Technologies (Ontologies, Semantic Web Services)</td>
</tr>
<tr>
<td>System Technologies (Data Management, System Maintenance)</td>
</tr>
</tbody>
</table>
1. Data management

- Data storage. This comprises aspects of data storage in a SW portal, especially data storage devices (database, RDF-Repository, etc.) and the kind of information that is stored (information-items, user-data, ontology-data, etc.).
- Sorting and indexing. Sorting and indexing techniques improve the system’s data storing and retrieving capabilities.
- Data transfer. This aspect comprises data formats and transfer protocols used in the system.

2. System maintenance

- System administration. Administration of a SW portal includes maintaining information items, user data and ontologies applied in the system as well as tool support for administrating the system at runtime.
- Security technology. Features for ensuring safety of information access should be inspected.

Semantic web technologies. Semantic web technologies to be utilized in a SW portal are ontologies and semantic web services.

Ontologies. Central components of a semantic web portal are ontologies. An ontology provides term definitions of the domain of interest and it can be applied in different ways to enable enhanced functionalities of a SW portal (Maedche et al., 2001). For describing the usage of ontologies in a SW portal is based on the following aspects:

- Ontology type. Different types of ontologies are distinguished according to the purpose of their usage.
- Ontology structure. An overview of the ontologies’ structure and size is given to achieve a basic understanding of the ontology used (Gomez-Perez and Benjamins, 1999).
- Additional facets. Optionally, further features of the ontology can be stated if they are important to understand this ontology usage in a SW portal. Example criteria are internationalization, multilingualism, and balance of expressivity and scalability of the ontology.

Inference and reasoning. Depending on the ontology formalism different mechanisms can be used to enhance systems usability. For example a reasoner can be employed to check cardinality constraints and class membership or an inference engine could interpret symmetric or transitive relationships.

Ontology management. The last aspect for evaluating ontology usage in SW Portals is ontology management, i.e. techniques for administrating ontologies. As the ontology is the central component to enable semantic web portals, appropriate ontology management facilities are essential for long-term usability of the portal. The aspects enumerated below are based on the requirements for ontology library systems identified in (Ding and Fensel, 2002). These requirements must not only hold for a single ontology, but also and especially for a network of multiple ontologies (Klein, 2002):

- Editing. An appropriate editing facility has to be provided. This can either be an ontology editor like PROTEGE[1], OntoEdit[2], or an editor facility integrated in the portal.
- Maintenance/versioning. An ontology is a static representation of knowledge structures. As the domain of a SW portal may change over time, the ontology should be updatable. This can be achieved by ontology versioning techniques which allow tracking changes by enumerating different ontology versions. Furthermore the system should provide means of keeping the relation between the schema and the instances consistent, in case of a change to the one or the other. For example support is needed to update instance data automatically in case a property is added to the corresponding instance, this can be for example achieved by sending an email to the owner of the specific instance data that needs to be updated.
Ontology search for administration. In order to facilitate management of several or huge ontologies in a SW portal, appropriate support for finding a specific ontology or a specific part of an ontology is necessary. This is different to the search functionality within the information life cycle (access), which is more end-user oriented and hides in particular the technical details of the ontologies which are necessary in the context of Administration.

Standardization/interoperability. In order to enable interoperability and information exchange with other SW portals and semantic web applications the ontology management system of a SW portal should support semantic web ontology languages and provide export/import functionalities for these. Syntactical interoperability is the first step towards a semantic one – enabling sharing of dynamically evolving ontologies in a peer-to-peer fashion.

Semantic web services. Web services add a new level of functionality on top of current web, transforming the web from a distributed source of information to a distributed source of functionality. Current web service technologies around UDDI[3], WSDL[4] and SOAP[5] provide very limited service automation support. In this context, web services are enriched using semantic information in order to allow automatic location, composition, invocation and interoperation, bringing the new concept of semantic web services (Fensel and Bussler, 2002). The use of web services and the use of semantic technologies to enhance these services must be evaluated as it reflects to what extent a given portal exposes its functionality as services accessible over the web. Thus, the following aspects are evaluated:

- Functionality. The different functionalities available on the SW portal, e.g. content search, content publication, etc., can be made accessible by using web service technologies. The degree of functionality exposition via web services determines to what extent a portal can be used not only through the user interface, but also programmatically. Thus, a comparison between the portal functionalities accessible through its user interface and the ones accessible through web services is related.

- Semantic web services. Enriching web services with semantic information allows automatic location, composition, invocation, and interoperation of services. Therefore not only the portal functionality exposed through web services must be considered, but also to what extent these services include automation support.

Information processing layer

Based on the evaluation of the grounding technologies layer this section exposes the evaluation criteria required for a functional analysis of the information processing features of a semantic web portal.

Information item workflow. As an analysis framework for describing the information processing layer we take the “document life cycle” as a basis. This has been defined for intranet document management systems and identifies five life cycle stages: creation, publication, organization, access and destruction/maintenance (Ginsburg, 1999). This model can easily be adapted for information items in SW portals. The aim of this analysis is to point out the usability of semantic web technologies in order to enhance information processing capabilities of the SW portals.

Collaboration features. These features facilitate the building of virtual groups – communities of interest and provide support for creating concrete output, such as information items that can be accessed by the community.
Information access layer

The evaluation criteria for information access layer are also important factors for SW portal as this layer is the front-end of SW portal for community user, which have the following focuses:

- **Usability**: It addresses the relationship between a portal and its users. For a SW portal to be effective it must allow users to accomplish their tasks in the satisfiable way.

- **General assessment as web technology**: As a SW portal is a special breed of web application, it should satisfy basic requirements usually used for the assessment of web sites: such as coverage, maturity of implementation, personalization and communication, reliability of information resources, help and documents.

3. Evaluation of semantic web portals

In our evaluation, 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 most successfully using these technologies and most closely fitting our definition of a SW portal. We have inspected two academic portals (Esperonto and OntoWeb[6] portals) and two commercial portal technology infrastructures (Empolis K42 and Mondeca ITM). Here we will provide a comparison between these portals following the three layers of our evaluation scheme. Other portals have been also identified, but they do not fit our definition of the SW portal, therefore, they are not included here.

Grounding technologies

Grounding technologies contain the key technologies to enable the basic function of the portal, which can be further divided into system technologies and semantic web technologies.

**System technologies.** Most of the portals take the traditional three-tier architecture: a database and/or a file system as backend for data storage layer, Java Servlet based user interface for the front-end, and various server components in the middle tier. For document storage OntoWeb leveraged existing document management framework functionality (ZOPE). Other evaluated portals just provide simple upload functionality and use the web servers’ file system. Data transfer has been achieved by either using existing protocols (such as JDBC, SOAP) or home-made solutions like passing serialized Java Objects directly over TCP/IP. None of them has adopted a fully service orientated architecture (SOA). Systems are administrated directly via various application servers and operating system mechanisms. Security for information communication is mostly achieved by providing password-protection for registered users or private areas, with methods offered by the employed application server.

**Semantic web technologies.** Semantic features provided by semantic web technologies are currently implemented in a limited way, such as providing taxonomy import and export features. The ontologies used in the portals are normally specifically developed for the according portal. The ontologies’ character is more static than dynamic and updates are only allowed to a limited extent. Inference or reasoning is limited to very low level, mostly restricted to simple inverse, transitive or symmetric properties of ontological concepts or relations.

The control of ontology data and information items is usually handled by different user levels. Normally these are portal administrator (full rights), registered portal members (some rights), and guest visitors (limited rights). Ontologies and instances are maintained separately update by using existing ontology editors such as Protégé2000, or home-made solutions such as WebODE ontology editor (Esperonto portal), OIModeler (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 for export of schema and instance data. Some heuristic rules have been added to achieve the consistency. None of them provide a sufficient versioning mechanism to trace changes between different versions of the ontology.
For internal representation for the ontologies, the academic portals mainly use RDF, the commercial products are mainly based on the topic maps paradigm. The expressiveness is usually equivalent with a taxonomy structure with relations. Although semantic web services are one of the unique and essential functions provided by the semantic web technologies, none of the portals implements or supports them. Future plans of some portals have been already made to further implement semantic web services function to their portals, but no concrete discussion on such issue is available and foreseeable.

Reviewing the state of the art in realization of grounding technologies layer in SW portals, one of the main challenges is to combine semantic web technologies and the already existing classical content and document management systems. Most of the portals have the focus on the one or the other: Esperonto and K42 clearly put the main emphasis on ontology management. Therefore these emphases result in a less user friendly environment for daily work but on the other hand offer well structured information items. ITM offers connectors to different CMS (although the have to be implemented by the developer of a portal application based on the ITM-technology), also putting the main emphasis on structuring and querying using ontology based techniques. OntoWeb is based on the ZOPE framework and provides a matured user interface in combination with a set of standard CMS functionalities.

Technically, the challenge is to reuse work out of both areas – the ontology and the document management area. Most approaches use a document as well as an ontology repository and they link data present in both systems, where the ontology repository handles most of the meta data management. Our challenge is to keep both repositories always synchronized; so, the harmonization and integration of semantic web technologies with existing content and document management systems and, furthermore, making them become an inseparable unit is a main task and challenge.

Information processing

The information processing facilities of an SW portal consist of five life cycle stages: creation, publication, organization, access and maintenance. Each portal varies in the implementation of these phases and borders are often intangible. In all portals investigated in detail, the creation of a new information item is based on HTML-forms which represent the attributes of the corresponding ontology concept. Usually the assignment of a new item to the ontology is achieved by this implicitly. For the creation of documents there is only limited support in SW portals. Most rely on external editors like Microsoft Word. K42 does not provide complete web based forms for end users. Esperonto portal provides this, but using the information item name as unique identifier, as well as problems to assign property values in one creation step. OntoWeb does provide complete forms and in addition integrates predefined 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 an Microsoft word document.

The publication of a new information item is usually divided into the submission by the creator and a validation by the portal administrator. The actual sub-steps in the publication phase vary between the portals, mainly depending on the number of different user levels. For access of the information by users, most of the portals provide ontology based navigation or browsing functions. Only ITM does combine this with a full text search of information item content. OntoWeb does not offer full text search and does not interweave the ontology with a thesaurus. K42 offers a wide range on visualization tools, but does not provide an ontological search form as defined in our criteria. Whereas Esperonto does, but due to less extensive usage of visualization (no graphical representation) and technical problems with their key word search the portals usability was reduced.

In the organization phase OntoWeb provides little functionality, since it maps its ontology to a simple object model (ZOPE objects), which does not support rich features like other ontology repositories. K42 and ITM both use self developed repositories, only Esperonto portal leverages the functionalities of a pre existing repository with a well developed set of functionalities like versioning, inferencing, ontology import and export, etc.
Maintenance is closely related to organization of the information items. Support for collaborative evolution of the ontology schema and easy modification of instance data should be provided. None of the portals had elaborated evolution concepts for the schema. Esperonto’s underlying ontology system does support versioning but this feature is currently not used within the portal. 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.

**Information access**

The academic portals are used as the document management and dissemination point for research projects; the commercial ones investigated aim at developers of web portals for different application areas, e.g. a conference portal or a portal solution for knowledge management (see demonstration sites of K42 and Mondeca ITM). The portals’ usability is mainly limited to creation and maintenance of information within application domains. Many other functions to facilitate community communication are ignored, such as discussion forum, mailing list archives or referential materials.

The content provided in the portals covers only special application domains without a broader extension to related areas and domains. Except OntoWeb, the portals do not allow personalization of information spaces. Ontology-based portals can easily keep high consistency in information access because instances are provided based on an ontology. Basic help and document are provided to facilitate the easy-use of the portal, but there exist much space to improve such functions, especially to guest visitors.

For navigation most approaches directly rely on the conceptual model, i.e. the underlying ontology. They render, for example, tree structures according to the conceptual model. Only SWWS takes a slightly different approach, separating both these aspects (navigation and conceptual model) by using an extra ontology for the visualization.

4. Related work

Semantic web is a recent and booming research area starting few years ago. Although increasing efforts have been devoted to surveying ontology-related research studies from various aspects, no survey can be found for evaluating SW portals. Here we present and compare some works done before which have various links to our work here.

Ying and Fensel (2002) conducted extensive survey on current existing ontology library systems. The coverage of this survey is very broad (including almost all the existing ontology library systems) and the focus is ontology management, which is also the key part of the SW portal. Based on this paper, we come out with the evaluation criteria for ontology management part for our portal survey. While our paper’s focus is evaluation of SW portal, many other functions besides ontology management have also been evaluated, such as information processing, information access and semantic web services.

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 comparing to our survey, this paper gives us good hint for designing our evaluation schema. We further extend their proposed generic framework for SW portal to include many other functions which we believe also the important features that successful SW portal should bear, such as functional ontology management (editing, browsing and
searching, versioning), semantic web services, ontology-powered searching, and information processing workflow.

So from our point of view, our survey on SW portals is quite unique with the respect to the survey coverage and the evaluation scheme provided.

5. Conclusion

The benefits of implementing these SW technologies can be easily identified or foreseen as semantic web technologies have the potential to increase the information consistency and the information processing quality of portals. The main benefit of all approaches is to be able to model a portal structure with an ontology. As shown in previous work, ontologies are suitable to represent consensus knowledge. Exactly that is needed to exchange information with a community of interest and enable automated processing of information items. Conventional portals try to tackle this problem with various structuring methods like content type, view, proprietary meta data elements, etc. But this often ends up in user confusion and incompatibility with other portals. There exist several methodologies to model ontologies which can be used to create a conceptual structure for a web portal in form of an ontology as the formal representation of a user consensus. The benefit of an SW portal is that it is able to load this initial ontology and build a system out of the box that can satisfy user needs. It will be custom tailored but still be standard compliant.

Reviewing the results in more detail we discovered that the current features available for ontology management have to be further improved; this even holds for the systems building on top of existing ontology infrastructure like Esperonto. Editing, versioning, search and interoperability should be further enhanced. No portal had a mature versioning concept, which deals with, for example, changes in the ontology schema and reflects it automatically on the instance level. Also – except Esperonto – most systems assume a single (static) ontology and do not consider interoperability issues between multiple ontologies. Most systems offer basic import and export functionality relaying on RDFS, OWL or topic maps; these standardization efforts should be further driven unifying to one standard like OWL.

Also the interoperability between different portals is not evolved. OntoWeb implements this by the RDFS exchange with OntoWebEdu and OntoWebRoadmap in the early stages, but assumes identical ontology structures and so far does not deal with the full heterogeneity of this aspect.

Looking more closely at classical content management facets we have seen that the reuse of mature technology – like OntoWeb does with the ZOPE framework – significantly improves the usability, reliability and scalability. That means a semantic web portal should not be developed from scratch, but reuse existing technologies where possible.

Another aspect which relates to our definition of a SW portal was also mainly neglected: community features that help building and tightening a group of interest were not implemented, only OntoWeb offers with the personal folder for registered users such a feature. Furthermore in general the implementations did not provide a community aware user interface (Grudin, 1994), for example, they take not into account that some users only occasionally use the portal and get confused by a too complicated interface.

Future development of successful SW portals should focus on not only the above criteria but also on semantic web services which will lift semantic web portals to next level. Semantic web services transform current web from a distributed source of information to a distributed source of functionality and a web portal is a descent platform to implement this. Various functions provided by current portals can be further refined as services. These services can be automatic located, composed, invoked and interoperaled with other services or agents available via the web, which can significantly extend the functionality of the portal.

Notes

1. see: http://protege.stanford.edu
2. see: www.ontoprise.de
4. www.uddi.org

5. www.w3.org/TR/soap/

6. We have evaluated the technology of the main portal: www.ontoweb.org/ However part of the OntoWeb Framework are also OntoWebEdu and OntoWeb RoadMap Portal, mentioned in section 4.5.

References


About the authors

Holger Lausen (1977) is a PhD student in DERI, University of Innsbruck, Austria. Holger obtained his diploma in Computer Science at Flensburg University of Applied Science in 2003. His diploma thesis discussed the integration of semantic web technologies and document management systems in a theoretical and practical manner. Since April 2003 he is active as researcher within DERI Austria. His research interests include semantic web and especially semantic web services. E-mail: holger.lausen@deri.org

Ying Ding is Assistant Professor at the University of Innsbruck. Previously she was a senior researcher at the Division of Mathematics & Computer Science at the Free University of Amsterdam. The concerns of her doctoral research are information retrieval, information visualization and knowledge management by using data mining techniques. She is very proactive in many consultancy projects between University and the companies. She has published around 40 papers in top-level journals and conferences. She is co-author of the book Intelligent Information Integration in B2B Electronic Commerce. Her current interest areas include semantic web, web services, semantic information retrieval, knowledge and content management, e-commerce and commercial application of semantic web technology. E-mail: ying.ding@deri.org

Michael Stollberg, MA (1977), is a PhD student in Computer Science at the University of Innsbruck, Austria. Before he joined the DERI Austria research group in May 2003 as a junior research member, he obtained his master degree in Communication Science, Computer Science, and Economics at the Free University of Berlin, Germany. His master thesis addressed the challenges on ontology technologies with regard to their usability as the semantic basis for the semantic web. Michael is involved in several research and IST-projects. In his research work he is concentrating on challenges of semantic web and
semantic web services technologies, with special regards to the resulting impacts on information system architectures. E-mail: Michael.stollberg@deri.org

Dieter Fensel (1960) works at the University of Innsbruck, Austria. His research interests include ontologies, the semantic web, web services, knowledge management, enterprise application integration, and electronic commerce. He is an editor of Knowledge and Information Systems: An International Journal (KAIS), IEEE Intelligent Systems, the Electronic Transactions on Artificial Intelligence (ETAI), and Web Intelligence and Agent Systems (WIAS). He has been involved in several national and internal research projects, for example, the IST projects COQ, Esperonto, H-Techsight, IBROW, Multiple, Ontoknowledge, Ontoweb, SWAP, SWWS, and Wonderweb. He has been the project coordinator of Ontoknowledge, Ontoweb, and SWWS. E-mail: dieter.fensel@deri.org

Rubén Lara (1979) is a PhD student in DERI, University of Innsbruck, Austria. Ruben obtained his MS in Computer Science at Universidad Autónoma de Madrid (UAM) in 2001, and received the First National Award in Computer Science by the Spanish Ministry of Culture and Science, as well as the Special Award in Computer Science at UAM. Since January 2003 he is a researcher on semantic web at DERI Austria, with special interest in semantic web services. E-mail: ruben.lara@deri.org

Sung-kook Han is a full professor in the Computer Engineering Department at Won Kwang University, South Korea. Prof. Han completed his PhD in Computer Science at the Inha University in Korea in 1988. His main research area is natural language understanding and knowledge representation to realize the intelligent systems. His current research focuses on semantic web services and ontologies based on conceptual processing of NLP and AI. His main research is to implement application systems using semantic web services and ontologies, as well as to research on the foundation theories for metadata model and the efficient inference mechanisms. Recently, he pursues semantic web service frameworks and practical knowledge-enhanced application systems that can show the potentiality of ontologies. E-mail: skhan@wonkwang.ac.kr