Organizational Semantic Web based Portals

Necula, Sabina-Cristiana

Alexandru Ioan Cuza University of Iasi

April 2011

ORGANIZATIONAL SEMANTIC WEB BASED PORTALS

Sabina Cristiana Necula

PhD Researcher
Research Department, Faculty of Economics and Business Administration, “Alexandru Ioan Cuza” University, Iaşi, Romania
Address: B330a, Bd. Carol I nr. 22, 700505, Iasi, Romania
Phone: ++ <40 232 201070>; Fax: ++ <40 232 217000>
E-mail: sabina.mihalache@gmail.com

This paper tries to treat organizational semantic web based portals. The first part of the paper focuses on concepts regarding semantic web based portals. After discussing some concepts we treat the basic functionalities that a semantic web based portal must have and we finish by presenting these functionalities by actual examples.

We present semantic web based portals after studying the necessary implementations from literature and practice. We develop some examples that use semantic web based technologies.

INTRODUCTION

Since the development of electronic computer people have wanted different types of automation and technologies and invented instruments that rely on primary technology: automatic data processing.

The evolution of systems was triggered by the need to integrate technologies due to the limits by which they were implemented. Even today technology providers promise to offer tools that handle useful information; although users still want current information. Concerns on enterprise modelling, enterprise architecture, and formal description of some informal data sources are the same: to provide information in the sense of the meaning perceived by the user.

But existing information technologies have limitations that have different descriptions of the data: manufacturers offer different ways for representation and organization of data.

We appreciate and support the so-called knowledge discovered in data not based on induction algorithms, but extracted from the user in real time and organized in ontologies as the basic principle of artificial intelligence: the separation of control over knowledge.

Information is synthesized by the use of indicators for measurement of components (factors) considered important for information models. Detection is done trough the importance of factors:

- Through observation (statistics);
- Based on common knowledge;
- Based on past experiences, which reveal important factors to take into account

General knowledge previously known and considered true for all cases relating to a business process, acts as a factor when a value of an indicator is in non-compliance with admitted values. The most important form of knowledge (the specialized, context-dependent, and actionable) depends on factors, indicators, plans, budgets and multi-dimensional simulation is like a combination of factors values.

To achieve the required strategic management views of the management information, means specifying a model for tracking the profitability of the organization. Planning and performance measurement are not confined to this level. Often involves an analysis of external data, an internal management accounting and external data.[Meier M., Sinzig W., Mertens P., 2005]

Even if the internal data sources are not a problem, often it happens that companies do not use to integrate management systems. Organizing information obtained by processing the data primarily require information integration.

Organizations develop physical and logical models of information system based on a function-oriented approach or process. Whatever the methodology chosen to develop and implement information systems, difficulties arise from poor organization (the real information system) and the limits of integration. Each participant in the development process has a different view of system models based on area of expertise to which it belongs. The user wants a system geared to the needs, the developer wants a system-oriented methodology that is used for the analysis phase, the implementer wants an adequate physical architecture available.

Inferring is closely linked to learning. In order to be retrieved information must be obtained in hard copy or as a result of inference. Learning is an important feature in ensuring the intelligent behavior of a computer model making.

Therefore, learning can be seen from two viewpoints:

- In terms of storage;
- In terms of the manifestation of the learning activity.

In terms of storage, information is the existence of a link between physical support structured data and their meaning.

The event detection means learning activity behavior data using intelligent technology or a previously specified behavior using conventional software.

User information depends on the quality and extent of information that is the decision-making. Currently, knowledge belongs to humans; the way they use information can be formalized only by explaining the rules. The rules are, in itself, the decision model. User interaction mode may be the solution either directly by providing decision or interactive questions and answers, but decision-making model is formalized in the previous IF ... THEN ... ELSE rules. There is currently no automatic way of achieving a link between a data structure in a certain context to the decision-makers' meaning (understanding), because now computers don't understand humans, unless this fact is previously specified.

In what follows, we present some methods for achieving these links automatically.

Google's executive director stated in an interview [Schmidt, E., 2007] that the strategy of indexing web pages is to count mouse clicks made by users of information in the Internet session. In other words, users are allowed to organize information according to their interest for a certain content. PageRanking Google search algorithm uses as criteria the number of visits and date of last update of the site.

Tough interesting, is the technology blog type with which users can edit the type hypertext links to other sources (other sites). Web 2.0 is a technology that wants to print a characteristic social web, and more importantly than that may represent a more efficient way of communicating than video conferencing, e-mail or chat for a server that stores information that members share to their community. Public and private information is not as happens with other types of technologies. Information is published by users of the community.

Creating links between information useful in decision making and decision can be achieved through the formalization of common knowledge. Although criticized in the literature this form of knowledge is suitable for situations or conversations that cannot be held to a blog or otherwise common knowledge representation is very useful in circumstances where is "without measure". CYC project initiated by researchers at Stanford University has proposed common knowledge representation. In our opinion, this project would have to be limited to human language formalization issues: the meaning of words antonyms, synonyms and homonyms. In other words, ought to have an orientation towards formalizing the meaning of a word in context. Yet researchers in this project formalize the domain knowledge (which is a form of explanation of knowledge).

Another very original way of creating a link between data structure and its meaning is training the users in the implementation of on-line. L. von Ahn has the sense of image capture technology for the user by using a game that gives users points when they assign a meaning identical images.[Ahn, von L., 20006]

Finding the necessary information based on content indexing is done by using search algorithms. Set indexing techniques applicable to the collaborative and integrative systems with which users can find information by specifying a keyword into a search engine. We can take such a collaborative application that stores the best practices for determining a problem or situation. Creating indexes for the files containing the best practices and key word search will be made by users.

We use the term Semantic Portal to refer to an information portal in which the information is acquired and published in semantic web format and in which the structure and domain model is made explicit (e.g. in the form of published ontologies).

There are several advantages to using semantic web standards for information portal design. These are summarized in Table 1.

Table 1 - Contrast semantic portals proposal with typical current approaches [www.w3.org]

Traditional design approach	Semantic Portal
Search by free text and stable classification hierarchy.	Multidimensional search by means of rich domain ontology.
Information organized by structured records, encourages top-down design and centralized maintenance.	Information semi-structured and extensible, allows for bottom-up evolution and decentralized updates.
Community can add information and annotations within the defined portal structure.	Communities can add new classification and organizational schemas and extend the information structure.
Portal content is stored and managed centrally.	Portal content is stored and managed by a decentralized web of supplying organizations and individuals. Multiple aggregations and views of the same data is possible.
Providers supply data to each portal separately through portal-specific forms. Each copy has to be maintained separately.	Providers publish data in reusable form that can be incorporated into multiple portals but updates remain under their control.
Portal aimed purely at human access. Separate mechanisms are needed when content is to be shared with a partner organization.	Information structure is directly machine accessible to facilitate cross-portal integration.

ONTOLOGIES

The use of explicit, shared domain ontology enables both data sharing and richer site structure and navigation including multidimensional classification and browsing schemes. Use of the Semantic Web standards for encoding these ontologies also enable the ontologies themselves to be shared and reused across portals. Several projects have already derived benefits from ontology-driven portal designs [SEAL][WEB-PORTALS].

EVOLUTION

Requirements change over time leading to extensions to the information model. The semantic web helps in two ways. Firstly, the user interface and submission tools can be generated from the declarative ontology. Secondly, the semi-structured data representation of RDF permits new data properties and types to be incrementally added without invalidating existing data, in such a way that both original and extended formats can be used interchangeably. This suggests an alternative approach to information portal design. Instead a long top-down design cycle, we start from a seed ontology and information structure that we extend incrementally.

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.

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 developed based on semantic web technologies.

The Esperonto Portal is a case study of the ODESeW knowledge portal generator developed by the Ontology Group at Facultad de Informática, Universidad Politécnica de Madrid. It serves as the intra- and extranet platform for the EU project Esperonto.8

The OntoWeb Portal 10 is a community portal for academic to industrial partners who share an interest in the Semantic Web. It was set up as part of the EU project OntoWeb (IST-2000-29243). It is built up on the ZOPE Application Server and a Content Management Framework (CMF) offered by the ZOPE Cooperation. 11 K42 is a knowledge management product developed by Empolis based on the Topic Map paradigm. It offers a basic infrastructure for storage, querying and maintenance whereupon portals can be realized by application developers. During our evaluation, the product portfolio within Empolis was changed: the K42 development as stand alone product is not continued but its functionality will be integrated in the e:kms knowledge suite.

Another portal is developed by iSOCO20 and serves as a dissemination platform for the EU-funded research project SWWS. It uses a domain ontology storing information about project partners, project members, the work plan with all the work-packages and all deliverables produced within the project. This ontology is created using Protégé200021 by the portal administrator. The ontology language used is RDF. Instances are also created using this tool and both the ontology and the instances are stored in files (there is no database support) and can be exported to the different ontology languages supported as by Protégé2000.

The AIFB23 at the University of Karlsruhe has build one of the first semantic web portal known to us. It was intended to be a platform for information exchange and collaboration for the “Knowledge Annotation Initiative of the Knowledge Acquisition community”. To structure knowledge, an ontology was developed as an act of international collaboration of researchers.

The ontology constituted the basis to annotate web documents of the knowledge acquisition community in order to enable intelligent access to these documents and to infer implicit knowledge from explicitly stated facts and rules from the ontology. This portal is no longer maintained and even if it is fully based on a ontology its functionality is rather simple (there was no web based administration functionality, for example) and the Information Access layer does not reflect modern user interface conventions and was hard to understand for new users.

Other portals were developed after KA2 was discontinued. Namely parts of the AIFB25 site are based on semantic web technologies and the Karlsruhe Ontology and Semantic Web Tool Suite (KAON26) includes a basic framework for portal creation27. Representative for this we evaluated the OntoWeb portal as the latest and most matured development of AIFB.

Semantic features as provided by unique Semantic Web technologies are currently implemented in a limited way, such as providing taxonomy import and export features. Reasons could be the immaturity of Semantic Web technologies and the difficulty of employing them due to technical reasons.

The ontologies used in the portals are normally specifically developed for the according portal, even though some of them are reusing existing ontologies (e.g. the OntoWeb ontology relies on the KA2 ontology. The ontologies’ character is more static then dynamic and updates are only allowed to a limited point; updates simply overwrite existing ontologies and very limited versioning mechanisms are used. None of the portal evaluated offers multi-language support for its ontologies. 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 by using existing ontology editors such as Protégé2000, or home-made solutions such as WebODE ontology editor (Esperonto portal), 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 for export of schema and instance data. Some heuristic rules have been added to achieve the consistency – such as when a concept is deleted from the ontology, its instance will become the instances of its super class. None of them provides a sufficient versioning mechanism to trace changes between different versions of the ontology. Also the support for matching instances to a changing ontology schema is very limited as outlined above.

CONCLUSIONS

The issues presented draw attention to the following issues:

- The factors that determine a key aspect in making decisions and conducting business processes are numerous (more than 2) so that the simulations must use models to forecast or multidimensional neural networks;
- The factors mentioned above are not only numbers but also quality and value of the type often contain more, less, very much, probably so. Simulations are unlikely in this case, without specifying an option earlier in the form of expert systems or knowledge of extracting knowledge about using real-time factors, the decision maker. It remains however a key issue to use this knowledge in an integrated information system with data processing;
- Source of input in determining the factors is different: often, it takes data from active markets, from internal management system or through surveys.

Aknowledgment

This work was supported by CNCSIS-UEFISCSU, project number PN II-RU code 188/2010.

References

1. *Meier, M., Sinzig, W., and Mertens, P. (2005), "Enterprise Management with SAP SEM™/ Business Analytics", second edition, Springer-Verlag, Berlin, p. 5*
2. *Schmidt, E. (2007), Eric Schmidt at Google Press Day Paris, The Official Google Channel, <http://video.google.com/videoplay?docid=-7725907465061773344&q=ceo+google+web&total=208&start=10&num=10&so=0&type=search&plindex=0>*
3. *Ahn, von L. (2006), Human Computation, <http://video.google.com/videoplay?docid=-8246463980976635143&q=human+computation&total=403&start=0&num=10&so=0&type=search&plindex=0>*
4. *http://www.w3.org/2001/sw/Europe/reports/demo_2_report/*
5. *Heflin, J. (2003), "Web Ontology Language (OWL). Use Cases and Requirements. W3C Working Draft", 31 March 2003*