INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 20 • 2011

ISSN 1507-3858

Ilona Pawełoszek
Częstochowa University of Technology

SEMANTICALLY ENHANCED INFORMATION PORTAL FOR COMMUNITY OF UNIVERSITY RESEARCHERS

Summary: The work of university researchers can be supported by using information portals based on open architectures. The portals can be enhanced and developed by exploiting Web Services, semantic technologies, mashup or microblogging technologies. The aim of this paper is to present selected tools and techniques for semantic enrichment of information portal that would serve community of university researchers, teams and students, to support their individual and collaborative work. First the paper addresses the general characteristics of the community software and Semantic Web. Next the required functionalities of the portal are presented followed by brief description of methodologies and technologies proposed for implementation of the functionalities in practice. Solutions are described on the example of the Czestochowa University of Technology.

Keywords: Semantic Web, semantic portal, Web 2.0, social networking, knowledge sharing.

1. Introduction

The need to manage information overload and accelerate communication inside an enterprise makes content management tools a crucial factor for every large organization. University faculties often have hundreds of workers in administrative or scientific positions, whose tasks are "knowledge intensive", that means communication, and information access are crucial for their work. Therefore information portals can effectively support and facilitate the different tasks of knowledge management at universities.

The concepts of enterprise information portals (EIP) and Knowledge Management have been considered simultaneously and jointly since 1998 as the way to improve enterprise-wide communication [Firestone 2003, p. 10]. Contemporary organizations seek to improve the use of tacit knowledge which is hard to capture and can be acquired only by socialization process. The arrival of the Web 2.0 paradigm in the year 2001 had a strong influence on all activities related to knowledge sharing. Social software that embraces wikis, blogs, social networks and bookmarks along with podcast and videocasts enable users to collectively build and develop new social forms of collaboration. Therefore Web 2.0. technologies are often incorporated into the field of KM as the tool especially invented for knowledge socialization.

The next step in the evolution of World Wide Web that promises great changes in the way people interact with the global network is Semantic Web sometimes related to as Web 3.0. Its aim is to make the data on the Web machine-interpretable so the computers could make use of the Web documents in intelligent ways like: extracting relevant data, answering the user's questions directly, reasoning on the bases of data found, searching for services, finding linkages between concepts, people and objects, providing the user with knowledge. The approach to building community portals also has changed since the emergence of Web 3.0. As the practice shows these two disparate trends may go on together well in supporting management initiatives.

Information portals are usually based open architectures and can be enhanced and developed by exploiting Web Services, semantic technologies, mashup technologies or microblogging tools. Universities are usually favorable environments to apply innovations especially in terms of information sharing and communication, because they employ open-minded knowledge workers, who often are innovators and researchers themselves. The aim of this paper is to present selected tools, techniques and guidelines for semantic enhancement of an information portal that would serve community of university workers, researchers, teams and students.

2. Web 2.0 and community software

The concept of "Web 2.0" began with a conference brainstorming session between O'Reilly and MediaLive International in the year 2001 [O'Reilly 2007, p. 17]. O'Reilly is a publishing company, whose mission is to spread the knowledge of innovators through books, online services, magazines, research, and conferences[1]. MediaLive International, on the other hand, works with companies as an IT and CRM consultant and also organizes tradeshows and conferences on information techno-logy[2].

There is no uniform, agreed definition of Web 2.0, since the term refers to many aspects of the functioning of World Wide Web, some of aspects and characteristics are enumerated in Table 1.

According to O'Reilly [2007, p. 10] Web 2.0 is a set of social, economic, and technology trends that collectively form the basis for the next generation of Internet – a more mature, distinct medium characterized by user participation, openness, and network effects. The definition proposed by O'Reilly seems to be comprehensive enough to describe technical and social impact of Web 2.0 on contemporary Internet.

All those aforementioned features are facilitated by exploitation of specialized software platforms, highly customizable and easy to use even for beginners. The best Web 2.0 software is inherently designed to harness collective intelligence through the architecture of participation [O'Reilly 2007, p. 13]. The term "collective

[1] http://oreilly.com/about.

[2] http://www.medialiveintl.com/.

Table 1. The characteristics of Web 2.0

Aspects of Web 2.0	Characteristics
Creativity of Internet users	• Active users' participation in Web content creation • Freedom of speech – lack of control or slight control • Creating online communities
Visual aspects of Web design	• Legibility of websites – large fonts • Gradient background and textures remainding surfaces of glass, metal or plastic • Possibility to browse websites on different devices • Simplicity in form, accessibility, personalization
Content of Web sites	• Mashup applications (composed of widgets from many different Internet sources) • Diversity of formats: texts, images, audio, video, interactive maps. • Personalization of information – adapting content to (known or expected) user's preferences
Internet technologies	• Ajax, XML, SOAP • Exposing Web APIs • RSS, Atom, Microformats
Content management systems	• Dissemination of easy content editing tools for creating Web sites without knowledge of HTML

Source: based on: [Pawełoszek-Korek 2008, p. 258].

intelligence" (CI) refers to the capacity of human communities to evolve towards higher order complexity and harmony, through such innovation mechanisms as differentiation and integration, competition and collaboration [Pór 2008, p. 238]. Put in another way CI may be seen as a common ability of a group of individuals to create knowledge. The most famous artifacts of CI in cyberspace are so called Wikis, with the most spectacular example: Wikipedia – Web-based encyclopedia co-created by thousands of editors worldwide.

Today's software platforms are designed with the architecture of participation in mind, this means they are meant for active user contribution in creation of content, functions and services. Web 2.0 platforms continuously grow in popularity due to the phenomenon of "network effect" which is driving force behind a number of users joining online communities. The network effect exists when the value of a good or a service increases along with the increasing number of people who use it. The online community or social network becomes the more interesting to a potential user the more resources it offers. On the other hand the amount of resources is dependent on number of users creating the given community. The commercial success of social networks is directly related to the number of participants. A remarkable feature of Web 2.0 is the possibility of creating new information value by content reusing, collaboration and information sharing.

3. The idea of Semantic Web

The next step in the evolution of Internet is to enhance the usefulness of Web platforms by harnessing machines intelligence. The aim of the new generation Web applications is to support users in many different tasks and make the interaction with data sources more effective. The next, more intelligent Internet iteration is known as Web 3.0 or Semantic Web. The new generation of Web, on the other hand, aims to make the data machine-interpretable so the applications could receive data, understand it and decide how to use it. From traditional search engine the user can only get information that is explicitly written somewhere in html documents or databases on the Web. Semantic Web applications are designed to work much like human brain, they have the ability to infer, so they can present the information that has never been explicitly written anywhere.

Certainly to achieve high level of intelligence in Web applications the approaches to design and development must change. One of the simplest attempts to make HTML more expressive in terms of semantics was the project of microformats. Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards[3]. Using microformats consists in adding special attributes (like rev, rel or class) with specific values to the HTML tags. The attributes express the meaning of data enclosed by the tag or relationship of some linked resource to the current document. Some examples of microformats are: XFN (relationships between people), geo and adr (location), hCard (contact information), hCallendar (event information), hReview and hResume (review and resume microformats), hAtom (syndicated content microformat) [Allsopp 2007].

Microformats are easy to use and read, but their semantics is strong enough only for simple solutions (such as: data exchange with calendar applications or e-mail clients). When it comes to more sophisticated applications with some features of intelligence, microformats are not expressive enough to codify knowledge.

Therefore the next iteration of HTML was the answer for these shortcomings. The eXtensible Markup Language (XML) was based on a concept of separation of what is displayed to the user from the other information meaningful to the computers. With XML it is possible to express (to some limited way) meaning of information published on a website, so it can be understood by applications. However, the possibility to define one's own hypertext elements made XML loose an ability of unambiguous interpretation. This meant that the program reading the website was not always able to know the meaning of the hypertext elements. For example in one situation a tag <title> could represent the title of a paper and in another the scientific title of a person. To identify the meaning of hypertext elements the World Wide Web Consortium proposed a Resource Description Framework (RDF), which aim was to add formal syntax for representation of knowledge [Ambroszkiewicz,

[3] http://microformats.org/about.

Mikułowski 2006, p. 102]. RDF (like all the other Semantic Web languages) is based on XML syntax, it uses so called triples written as XML tags to express information as a sentence. So called RDF triples can be represented as a file or graphically, as a directed graph. Graphs can be linked to create a larger semantic network describing relations in a given domain (Figure 1).

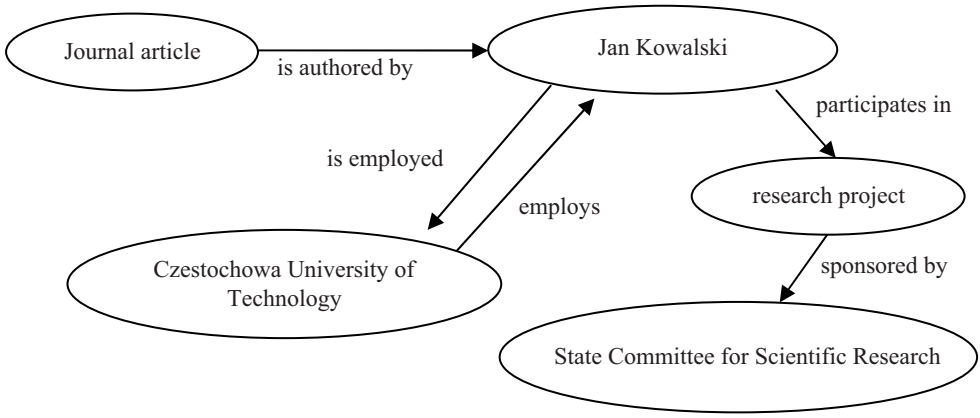

Fig. 1. A fragment of semantic network

Source: own study.

To decide whether the relations coded in RDF file or graph are consistent there is a need for specifying a number of relations that are possible and permitted for a given domain. The set of facts, relations and hierarchy between concepts in a given domain is referred to as ontology. There are special languages for defining ontologies – RDFS (RDF – Schema) and Web Ontology Language (OWL), both based on XML and RDF syntax. Different languages offer various levels of semantics, the RDFS is the simplest form of knowledge description. OWL has three increasingly expressive variants: OWL Lite, DL and Full. Each of them can be used depending on the different degrees of semantics to be achieved. There are four basic kinds of representation that can be used to model and semantically organize concepts, that is, controlled vocabularies, taxonomies, thesaurus, and ontologies.

RDFS and OWL Lite provide enough expressiveness for constructing thesauri and other taxonomies. They also have lower formal complexity than OWL DL and OWL Full, so it is easier to construct applications that will use them for reasoning. OWL DL allows constructing compound ontologies, it gives maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). OWL Full gives maximum expressiveness, the syntactic freedom of RDF with no computational guarantees.

Advanced Semantic Web applications may use selected languages along with reasoning mechanisms (such as Pellet, Kaon2, Racer, FaCT). For simple solutions it is enough to employ RDF databases and specialized query languages (such as SPARQL or RDQL), similar to those used in Relational Database Management Systems. One of the most popular toolkits for Semantic Web solutions is RAP (RDF API for PHP). It offers features for parsing, manipulating, storing, querying RDF data.

Aforementioned knowledge description tools along with web programming languages like PHP or ASP may be used to enhance Web portals with semantic features.

4. Guideliness for semantically enhanced information portal for Czestochowa University of Technology

Information portals allow easy access to heterogeneous data resources, applications and services in a consistent way. Portals may also improve corporate communication by exploiting various tools based on Web 2.0 paradigm. Standard portal features may be improved employing semantic technologies. A semantically enhanced information portal for a community may be understood as a user-centered web application exploiting selected Semantic Web technologies.

This paper proposes guidelines to create a single point of access to data and applications for a community of university researchers at Czestochowa University of Technology. The considered community consists of over 1000 people who work at the university as researchers, practice as educators and also take positions in the administrative hierarchy of the University. Students are also important part of the community, they need information to learn, take part in projects, conduct their own research for master or bachelor degree. Therefore students also should be taken into account while designing information portal.

The professional and scientific practice of the community members include:

– taking part in projects as a member of the team or leader,
– conducting individual research,
– participating at conferences,
– authoring and co-authoring of research papers, books, chapters, etc.
– editorialship of scientific publications,
– teaching,
– being in charge of organizational units (faculties, departments, chairs), managing subordinates.

The profile of Czestochowa University of Technology is very broad (the University consists of 6 faculties). The presentation in hereby paper will be narrowed down to management faculty. The faculties focus on different domains but their organizational structure is similar, so such simplification is acceptable in a model approach. At present a portal of the faculty is built in a traditional way, this means it does not allow user contribution in content creation and it offers very limited search

possibilities. It is only possible to search for worker's name and surname or browse the organizational structure of the faculty. All the information about workers are input by a webmaster and every time there is a need for changes or updates it requires his intervention, thus the information about workers' is very limited. A proposition of a new semantically enhanced portal is based on the following requirements:

– users can promote their work and publications, share interests among the community members,
– the portal helps to share detailed information about research projects and upcoming conferences,
– the portal facilitates searching for cooperatives according to many criteria (such as: domains of interests, realized projects, position in organizational structure of the faculty),
– active user contribution (every user can update his/her profile at any time and decide what information will be publicly accessible).
– the information from portal may be easily imported to external applications if needed.

To provide users with such functionality portal deals with following technical problems:

– creating community-specific ontology,
– choosing the right software platform,
– building an RDF database describing entities (workers, organizational structure, projects, publications, conferences)
– providing intuitive user interface for editing profiles and searching,
– exposing information for external applications.

The first step in creating semantic applications is to identify the concepts and relations in a domain under consideration and building ontology. It is probably the most tedious task that requires skills in ontology construction, proper tools, domain knowledge as well as identifying the community needs. The ontology may include definitions of terms (such as "publication", "journal paper", "author", "editor", etc.) and relations between terms, for example: "all journal papers are publications" or "all publications have one or more authors". In case of the portal for community of researchers fulfilling aforementioned requirements of usability and functionality it is necessary to use many ontologies, such as:

– description of persons containing: place and hours of duty, interests, curriculum vitae, participation in projects, publications, co-workers, position in university hierarchy, membership in organizations,
– organizational hierarchy of the University: names of positions, organizational units, responsibilities.
– projects ontology embracing: project status (not yet begun, looking for participants, during realization, finished), domain of project, sponsors, etc.
– conferences ontology: place, and time, deadline for submission, conference Web site, etc.

– ontology of publications: kind and rank of publication, language, hyperlink to library resources where the publication is available.
– ontology of research domains and interests.

Many of the ontologies mentioned above already exist and are free to use, although it may require some adjustments. Thanks to the specifics of open standards like OWL and RDF, ontologies can be extended or merged together if needed to fit special needs.

FOAF (Friend of a Friend) ontology is one of the most popular RDF (Resource Description Framework) vocabularies in the World Wide Web, it describes the basic information about people and their relations so it imposes an ideal solution in the considered case. Many useful ontologies (such as projects, conferences, publications, research) can be found at UMBC ebiquity portal[4], or by Semantic Web search engine like Swoogle[5]. Ontologies can be manipulated using many application tools, such as: OilEd, OntoEdit, Protégé-2000 or WebODE, to name a few. Some of them are distributed as Open Source software or are free to use for noncommercial purposes.

The most popular platforms for developing Semantic Web applications are: Jena[6], RAP (RDF API for PHP)[7] and ARC[8]. The first one is based on Java programming language and the former two are PHP classes. We propose to use RAP or ARC because of existing server platform used at Management Faculty (Apache Server with MySQL 5 and PHP 5.0). Using PHP libraries is straightforward – it does not require installation of any additional software components or server reconfiguration. PHP libraries and new functions can be seamlessly added to existing portal solution. It will be only necessary to add a database for storing RDF files with the structure adhering to the solution at choice (ARC or RAP).

The next important issue determining the success of a community portal solution is an intuitive and friendly user interface, considering that user's contribution is the basic assumption. The user interface should look like a conventional content management system with Web forms. It is strongly recommended to use techniques like Ajax and Javascript, first to prompt the values that are available and correct for every field and secondly to validate user input before sending data to the server. The values from the form are passed to the PHP script which converts them to RDF format and puts them to the RDF database.

Another important aspect of the portal functionality is to provide users with easy to use and intuitive searching mechanism. Searching RDF databases is realized by means of special query languages like SPARQL or RDQL, so the user input should be translated by a script to the appropriate syntax of a chosen query language. Search-

[4] http://ebiquity.umbc.edu/ontology/.
[5] http://swoogle.umbc.edu/.
[6] http://jena.sourceforge.net.
[7] http://www4.wiwiss.fu-berlin.de/bizer/rdfapi/index.html.
[8] http://arc.semsol.org/.

ing tool should support traditional keyword based search as well as the possibility to narrow the results to a given category (a concept from ontology). For example let us assume the user inputs a surname "Kowalski" as a keyword and narrows the results to the "projects" category. As an answer the system should display hyperlinks to all the projects the person of a name "Kowalski" is involved in. The user may want to find all the people who could co-work in her research project and are interested in informatics. The semantic search engine can conclude on the bases of domain ontology that if a person can create databases, design computer graphics or program in C++ he or she is also interested in informatics (although such information is not directly written in the person's profile).

The search results may be presented in many ways. For Semantic Web search engines it becomes more and more popular to present information in nonlinear and visually attractive way – as a graph, the edges of which show the relations between concepts or documents and the nodes represent documents themselves. The pictograms representing nodes are clickable and constitute hyperlinks to Web resources related to the given term.

Some features of a semantically enhanced portal may be exposed publicly to be used in external applications. It could be convenient for workers to subscribe information about updates in projects or new conferences using RSS/Atom to have the news right at their e-mail client or Web browser. Some information about workers, such as hours of duty could be available as XML Web Services so they may be imported to an information portal for students or IT systems of administration. Information from portal could also be used to provide an insight into activities of the community members. It would be much easier to draw up reports required by accreditation board and for statistical purposes.

The next interesting application enhancing communication abilities of the community is microblog. Microblogging have recently become a "must have" part of many community portals (i.e. Facebook). The idea behind microblogging is enabling users to send and read short messages (up to 140 characters) from other users. As the practice shows, microblogging is very useful and effective communication channel in many organizations [Zhang et al. 2010], as well as the tool of public relations.

One of the most important issues determining the success of every, either desktop or network, application is its user interface. Many studies prove that even the most functional application may be completely useless and disparaged by users if it does not offer reliable, intuitive and user-friendly interface. Apart from those three features, recently, a very important issue is personalization – understood as an ability to adapt the interface to user's individual needs. A mashup technology has been developed especially for building personalized interfaces. The semantic portal for community of university researchers can also use mashup for personalization purposes. Every user (researcher, teacher or student) has different information needs, thanks to mashup technology she can choose what kind of information should be at first shown up in her browser. Some examples of widgets are: the newest books in

library from the category of interest, microblogging channels of coworkers, teams, students groups, incoming conferences and seminaries on a given subject, local news from university life, interesting lectures, sport events, news about scientific or business research projects, important announcements for university students or workers, European programs and projects.

When the user enters the portal and logs in, it displays the static content and the set of selected widgets in places chosen by the user. Using mashup technology makes the portal solution highly customizable and open for further development.

5. Conclusions

Many discussions and interviews with community members at Management Faculty reveal their unsatisfied information needs, and necessity for robust knowledge sharing and communication tools. A remarkable improvement can be achieved by harnessing together Semantic Web and social networking tools. Probably the most expensive part of the discussed undertaking is designing and building ontologies according to the identified needs of community members. But the expenditure may turn out to be worthwhile if the portal brings results in facilitating collaboration and information access. It is hard to evaluate the success of the semantically enhanced portal in a quantitative way. Some measures can be applied, like: number of new projects, number of emerging research teams, number of publications, especially co-authored by groups of portal users. Future research directions can encompass distributed semantic information portals connecting different communities from other research and education centers in Poland.

References

Allsopp J., *Microformats. Empowering Your Markup for Web 2.0*, Apress, Berkeley, CA, 2007.

Ambroszkiewicz S., Mikułowski D., *Web serwisy i Semantic Web. Idee i technologie*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2006.

Firestone J.M., *Enterprise Information Portals and Knowledge Management*, KMCI Press/Butterworth Heinemann, Boston, MA, 2003.

O'Reilly T., *What is Web 2.0: Design patterns and business models for the next generation of software*, "International Journal of Digital Economics" 2007, vol. 65, pp. 17-37, http://mpra.ub.uni-muenchen.de/4580/1/MPRA_paper_4580.pdf (July 2010).

Pawełoszek-Korek I., *Zarządzanie wiedzą organizacji w świetle trendu Web 2.0*, [in:] H. Howaniec, W. Waszkielewicz (eds.), *Współczesne problemy zarządzania przedsiębiorstwami w gospodarce rynkowej*, Wydawnictwo Akademii Techniczno-Humanistycznej, Bielsko-Biała 2008, pp. 254-264.

Pór G., *Cultivating collective intelligence: A core leadership competence in a complex world*, [in:] M. Tovey (ed.), *Collective Intelligence: Creating a Prosperous World at Peace*, Earth Intelligence Network (EIN), Oakton, VA, 2008, www.oss.net/CIB (July 2010).

Zhang J., Qu Y., Cody J., Wu Y., *A case study of micro-blogging in the enterprise: Use, value, and related issues*, Proceedings of the 28th International Conference on Human Factors in Computing Systems 2010, ACM Press 2010.

SEMANTYCZNIE WZBOGACONY PORTAL INFORMACYJNY DLA SPOŁECZNOŚCI PRACOWNIKÓW NAUKOWYCH UNIWERSYTETU

Streszczenie: Praca naukowców zatrudnionych w szkołach wyższych wymaga ciągłego zdobywania wiedzy, dzielenia się nią, współpracy w zespołach. Współcześnie zadania te mogą być wspomagane poprzez zastosowanie portali informacyjnych. Celem tego artykułu jest prezentacja wybranych narzędzi i technik pozwalających wzbogacić semantycznie portal służący społeczności uniwersyteckiej do dzielenia się wiedzą. Proponowane rozwiązania są rozważane na przykładzie Wydziału Zarządzania Politechniki Częstochowskiej.

Słowa kluczowe: sieć semantyczna, portal semantyczny, Web 2.0, nawiązywanie społecznych kontaktów, dzielenie się wiedzą.