INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 1(31) • 2014

ISSN 1507-3858

Ilona Pawełoszek
Politechnika Częstochowska

SEMANTICIZING INNOVATIVE KNOWLEDGE

Summary: The paper addresses one of the technological aspects of knowledge creation for innovations – namely supporting knowledge externalization and sharing in an organization by using semantic Web solutions. Specifically, we propose a semantic approach to preserve and leverage knowledge from individuals to the organization. Thus, the organization will be able to continually learn and improve based on its employees' experience. We propose a solution based on the Semantic MediaWiki software platform, that would serve as a semantic middle layer facilitating access to innovative ideas and dispersed information resources. The presented case study from metallurgical industry aims to highlight the possibilities of the semantic integration of different data sources by using enterprise taxonomies to support innovative knowledge sharing among a community of knowledge workers.

Keywords: semantic MediaWiki, enterprise taxonomy, innovative knowledge, semantic structure, semanticization.

DOI: 10.15611/ie.2014.1.05

1. Introduction

Knowledge creation and innovation are two different processes that have a strong but complex relationship. Continuous innovation requires a well-planned system of knowledge management and an appropriate organizational environment that enables a company to excel in technological, market and administrative knowledge creation [Popadiuk, Choo 2006, p. 302]. Both the literature on knowledge management and that on innovation highlight the direct connection of these two domains [Jurczyk-Bunkowska, Chwastyk 2013, p. 43; O'Looney 2002, p. 122]. In particular, the authors emphasize the importance of the knowledge creation process which acts as a key enabler for the innovation process [Esterhuizen et al. 2012].

Many organizations have realized that a large amount of problems arise from un-captured or unshared knowledge. Specifically, they have discovered the need of knowing 'who knows what' [Lindvall et al. 2003] and the need to "know what they already know" [Stewart, Ruckdeschel 1998; Tiwana 2000, Weijermars 2011 p. 55].

A good set of knowledge management tools and practices prevents organizations from constantly "reinventing the wheels" and therefore powers-up the innovation creation process.

In this paper we address one of the technological aspects of knowledge creation for innovation – supporting knowledge sharing by using semantic Web solutions. We specifically propose a semantic approach to preserve and leverage tacit knowledge from individuals to the organization. Thus the organization will be able to continually learn and improve based on its employees experience.

2. Semantic Web Technology for supporting knowledge creation and innovation

Contemporary content management tools must effectively deal with the breadth, depth and complexity of information in terms of publishing, searching for and presenting data in a user-friendly manner. The key technologies are web-based communication and collaboration platforms for Internet and intranet usage. The organizational application of intranet technology tends to evolve and increase in sophistication over time [Damsgaard, Scheepers 2001, p. 676], however some studies reveal that the biggest problem with enterprise intranets today is the huge difference in user experience when compared to the Internet [Mukherjee, Mao 2004; Frischmuth et al. 2012, p. 10].

Moreover, the intranet should constitute a single point knowledge shop for quality information with certainty, authority and consistency [Cui et. al. 1999, p. 103]. Today's software platforms are designed with the architecture of participation in mind, this means they are meant for active user contribution in the creation of content, functions and services. Web 2.0 platforms are continuously growing in popularity due to the phenomenon of the "network effect" which is the driving force behind a number of users joining online communities. The same ideas should apply to intranets and corporate portals. For effective KM across the organization, we need the tools that facilitate sharing such as:

- Knowledge Portal – enabling the user to have one, simple way of navigating towards the desired information, supporting communication and collaboration.
- Users Knowledge Profiles – describing each user, created and maintained by the users themselves. Users' profiles are the answer to the need of knowing 'who knows what' in the organization.
- Collaborative workspace – a shared electronic area, where group members can work collaboratively.
- Discussion forum – within the intranet it is very useful to have a facility to pose any ad hoc question and receive back any responses from across the organization.
- Document Libraries are contextually organized collections of files that can be shared among the users – i.e. a library of common documents for a project.

- Knowledge repositories – contain both structured and unstructured information valuable for users: memos, notes, links to discussion forums, blog posts, document templates, multimedia files, reports and case studies etc.

All of the aforementioned tools and information resources may be potentially very useful. Semantic Web technology could be the key linkage between various "knowledge islands". With semantic annotation and search tools it is possible to link the documents according to their subjects, regardless of their provenance or place of storage. In the next section we propose a semantic solution based on the MediaWiki software platform, aiming to resolve the problem of "knowledge islands".

3. The Specifics of Semantic MediaWiki as a Collaborative Knowledge Sharing Environment

The aim of wiki is to organize and share information resources in a collaborative environment. Collaborative authoring is the effective way of creating knowledge, proved by many examples, of which the most spectacular is Wikipedia. The concept of wiki was coined in 1994 by [Cunningham 2006]. According to Cunningham "wiki is created by community" and it is designed in a way that allows everybody to create and edit whatever and whenever one wants if he/she finds the legacy content incorrect or incomplete [Richardson 2006].

The wiki technology is increasingly being used in corporate environments to facilitate a broad range of tasks [Lykourentzou et al. 2012] since it supports the creation of large and well-structured websites for various purposes including individual note-taking, managing documentation of cases, bibliographies, analyzing data and many more [Schneider, DaCosta 2014] For example, Motorola has more than 3,300 wikis that are used by 69,000 employees and further 8,000 business partners. The total wiki content measures 5 TB of data and covers a wide range of topics [Reynolds 2010, p. 166].

Despite the obvious benefit of a wiki, its contents are barely machine interpretable. Structural knowledge about how concepts are interrelated can neither be formally stated nor automatically processed. Numerical data is available only as plain text and thus cannot be processed considering its actual meaning. Since wikis are by nature unstructured, the use of semantics can greatly facilitate the usage of information and retrieving knowledge [Uden 2011, p. 235].

The most popular wiki engine – MediaWiki – is a free software open source package [MediaWiki 2014]. The basic idea behind it is that the programmers can read, redistribute, and modify the source code free of charge. The software is then improved much faster than would otherwise be the case [Epstein, Politano 2002]. There are hundreds of plugins (extensions) that power the core MediaWiki engine and allow to customize it for many purposes. One of these extensions is the Semantic Mediawiki package.

The idea behind the semantic Web technology is to provide meaning to the words and sentences published on the HTML pages so that their content can be understood and processed by machines. The understanding of data enables the machines to combine, compare and analyze them. Semantic Mediawiki (SMW) is a package of extensions to the MediaWiki platform that allows for the practical realization of the assumptions of the Semantic Web. The basic advantages of using the SMW platform are:

- automatic lists generation,
- improved data structure,
- advanced search capabilities,
- greater coherence of different language versions,
- possibility to export data in formats such as: CSV, JSON and RDF,
- possibility of exploiting Wikipedia and other wiki services as sources of knowledge.

The wiki software is created to serve as a platform for online encyclopedia, so each wiki page should describe an object of a particular category, a category, or an attribute of a category. The first impression on SMW pages is that they do not differ at all from standard wikis – they are human-readable. The difference lies in the code of the page which contains annotations defined in the special semantic wiki syntax. The annotations inform the wiki semantic search engine about the meaning of the elements (words, numbers, pictures), assign a page to a given category and define attributes for that category. Every SMW page describes an instance of some class or a definition of the class. In SMW terminology the "category" can be understood as an equivalent to the "class" in object programming.

The users can easily create their own categories with attributes at any time and this does not influence the existing resources. Adding annotations to a text is an easy task based on easy syntax similar to MediaWiki hyperlinks [Semantic Mediawiki 2014]:

[[Attribute name:: attribute value]]

In the above syntax only a text after two colons is displayed on the wiki page and it looks like a hyperlink. That text is understood as a value of an attribute, clicking on it moves the user to the page where the specified attribute value is defined.

One of the advantages is the fact that to create semantic resources in SMW it is not necessary to use ontology editing tools, nor to have the skills of semantic Web languages like OWL or RDF. The aforementioned features make the SMW an ideal platform for the collaborative creation of semantic resources by a community of non-technical users. Bearing in mind the information needs of the users along with the specifics of the MediaWiki software, we decided to build a solution that would serve as a semantic middle layer facilitating access to innovative ideas and dispersed information resources. A simplified idea of the proposed solution is shown in Figure 1. The dashed line encompasses the internal resources accessible via intranet.

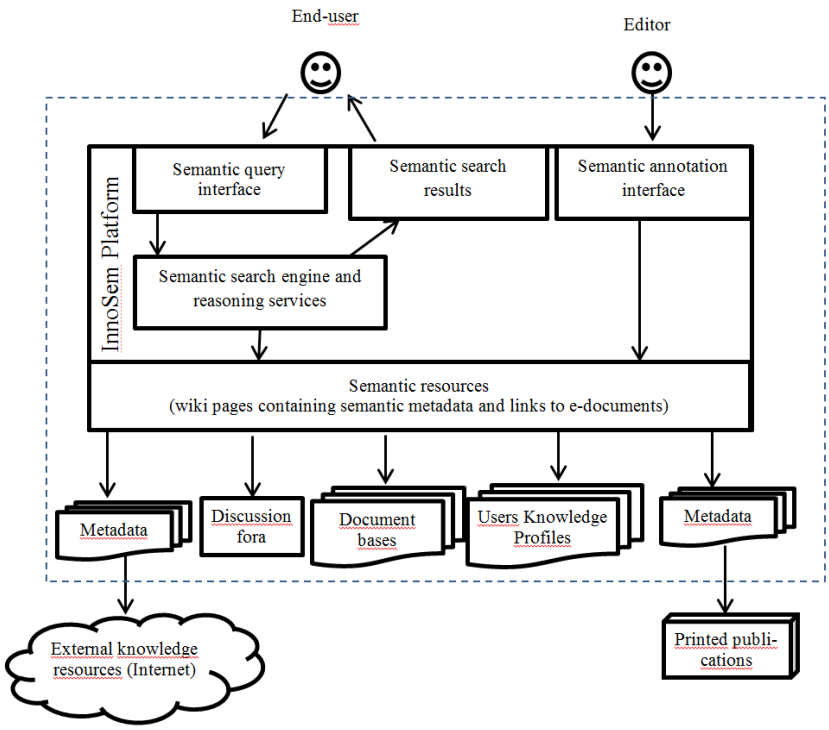

Fig. 1. The concept of InnoSem platform

Source: own study.

The semantic resources are:

1) pages containing semantic metadata descriptions of external information and knowledge resources,

2) pages containing the semantic metadata describing printed publications available on the bookshelves of various departments,

3) pages containing semantic metadata related to internally accessible documents that cannot be directly annotated (i.e. pdf files),

4) users profiles, notes and memos written directly in the wiki, and semantically annotated.

4. The Semantic Structure of the InnoSem Platform

The semantic structure defines the categories, subcategories and attributes of the entities in the system. The attributes are also categories. The semantic structure of the knowledge base for supporting innovations should be aligned to the specific of the given organization and its activities.

The semanticization of information resources is the process of defining and applying a semantic structure to a set of documents, it is also a form of knowledge formalization because the semantically annotated resources can be interpreted and processed by machines.

We propose the prototype semantic model of a digital document base of a medium sized metallurgical company. The company designs and produces a wide range of metal components mainly for industrial purposes, in small as well as in large-scale volumes. Due to many non-standard customer requests, the manufacturing processes must be highly flexible and require an innovative approach. The company has found that knowledge transfer and preservation do not function optimally and that significant chunks of knowledge are lost after the completion of customers' orders. Operation sheets and final production reports, intended to transfer and save the experience acquired, typically contain only a fraction of the effectively accumulated knowledge. Therefore subsequent manufacturing tasks rarely profit from this empirical knowledge and frequently the result is duplicated work, repeated mistakes and delays of order fulfillment.

We propose the lightweight semantic integration of document bases using Semantic MediaWiki to store and semantically annotate the innovative knowledge assets. The task of bringing all the documents to the MediaWiki platform and adding annotations of wiki-text is seemingly simple, although it requires a lot of manual work. The artefacts that cannot be directly moved to MediaWiki (i.e. multimedia files, pieces of software) are represented by metadata describing their content along with the link to the original asset.

However the critical task in the undertaking of semanticization is developing the right semantic structure which would fulfill the needs of the wiki users. The most important part of the semantic structure are enterprise taxonomies, since these taxonomies already reflect a large part of the domain terminology and corporate and organizational culture [Frischmuth et al. 2012].

In this case study the enterprise taxonomies include:

- Functional taxonomy – represents the business model and organizes information around the services and/or functions the company performs. The taxonomy is organized by customers and orders, the data source for constructing this taxonomy is the CRM system database.
- Organizational taxonomy reflects the structure and functions of production and R&D and also describes the workers' profiles structure.
- Data source taxonomy attempts to categorize and label the documents by the nature of the content, examples include: reports, operational sheets, policies and procedures, manuals, images, notes.
- Product taxonomy – defines the categories of final products and parts.
- Material taxonomy – defines the categories of materials used for production and their properties.

- Equipment Taxonomy – describes the types of tools and machinery, their properties such as: purpose, effectiveness, physical location.

The above taxonomies define terminology and provide a coherent description and shared conceptual structure of information resources for all the knowledge workers concerned with the creation, evaluation and practical application of innovative knowledge. The taxonomies enable annotating information coming from different sources in an explicit and unambiguous way. Figure 2 depicts some of the concepts described by taxonomies.

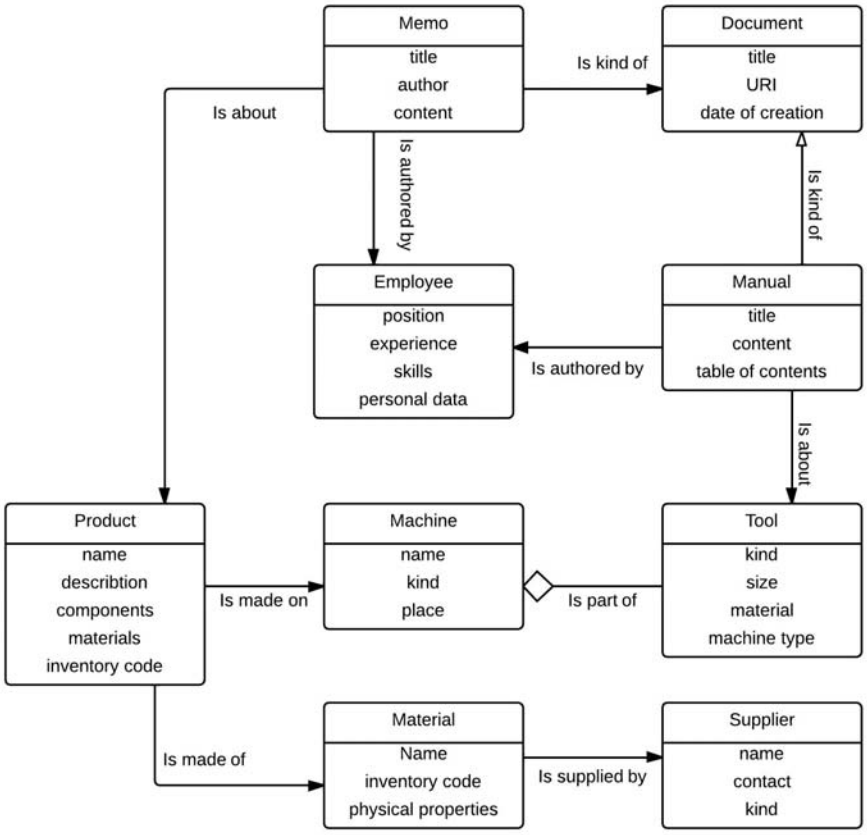

Fig. 2. A piece of semantic structure of a prototype InnoSem system

Source: own study.

Considering the presented fragment of semantic structure, it is possible to pose compound queries, for example a query to find all the memos written by workers that include information about products that are made of materials from supplier: Company X.

In SMW syntax the query would be:

[[Category:Memo]] [[is authored by::Employee]] [[is about.is made of. is supplied by::Company X]]

Moreover, in semantic SMW it is possible to define inverted properties and relations, so it would be also possible to find, for example, all the suppliers of a given material used in the production of the product Y, and the query would be:

[[Category:Supplier]] [[-is supplied by.-is made of::Product Y]]

In a traditional keyword based search it would be impossible to pose such a detailed query. As we can see in the example above, the wiki syntax is simple and similar to natural language sentences. In order to enhance the capabilities of posing compound queries we need to build enterprise knowledge bases specifying attributes of all the taxonomy classes and the relations between classes. The relations are also specific kind of attribute. To build a comprehensive knowledge structure it is critical to use shared concepts and terminology while defining classes, attributes and relations. Subsequently the emerging enterprise knowledge structure can be used for interlinking and annotating content in enterprise semantic wiki.

5. Conclusions

Innovative knowledge management requires the continuous dynamic process of development and adjustments to the existing knowledge structure. The approach presented with the case study from metallurgical industry aims to highlight the possibilities of the semantic integration of different data sources to support innovative knowledge sharing among the community of knowledge workers. Knowledge management can be supported by a variety of tools – from simple useable checklists to complex context-sensitive and artificial intelligence-based systems.

The solution based on the Open Source MediaWiki platform proposed here is easy to learn and maintain even by non-IT users, and whilst at the same time offering quite advanced semantic searching capabilities. The InnoSem prototype is in an early phase of its practical realization; the next research direction that will be described in the author's forthcoming publications will consider the issues of user experience assessment and the possibilities to customize the interface to suit the users' own ways of working.

References

Cui Z., Tamma V.A.M., Bellifemine F., 1999, *Ontology management in enterprises*, "BT Technology Journal", vol. 17, no. 4, October, pp. 98-107.

Cunningham W., 2006, *Design principles of Wiki: How can so little do so much?*, Retrieved 01.2014, from: http://c2.com/doc/wikisym/WikiSym2006.pdf.

Damsgaard J., Scheepers R., 2001, *Using Intranet Technology to Foster Organizational Knowledge Creation*, Global Co-Operation in the New Millennium Proceedings of the 9th European Conference on Information Systems Bled, Slovenia, June 27-29.

Epstein M.A., Politano F.L., 2002, *Drafting License Agreements*, Fourth Edition, Aspen Publishers, pp.13-46.

Esterhuizen D., Schutte C., Du Toit A., 2012, *Knowledge creation processes as critical enablers for innovation*, "International Journal of Information Management", vol. 32, Issue 4, August 2012, pp. 354-364.

Frischmuth P., Klímek J., Auer S., Tramp S., Unbehauen J., Holzweißig K., Marquardt C.M., *Linked Data in Enterprise Information Integration*, Semantic Web 2012, https://www.academia.edu/2823777/Linked_Data_in_Enterprise_Information_Integration.

Jurczyk-Bunkowska M., Chwastyk P., 2013, *Planowanie procesów innowacji. Aspekty teoretyczne ilustrowane praktycznymi przykładami*, Politechnika Opolska, Opole.

Katz B., 2007, *The integration of project management processes with a methodology to manage a radical innovation project*, MSc thesis, Dept. of Ind. Eng., Stellenbosch Univ., South Africa, http://hdl.handle.net/10019.1/2065.

Lindvall M., Rus I., Sinha S.S., 2003, *Technology Support for Knowledge Management*, "Technology Support for Knowledge Management. Lecture Notes in Computer Science", vol. 2640, November.

Lykourentzou I., Dagka F., Papadaki K., Lepouras G., Vassilakis C., 2012, *Wikis in enterprise settings: A survey*, Retrieved (May 2014), http://www.sdbs.uop.gr/files/wiki-technical-report.pdf

MediaWiki, 2014, http://www.mediawiki.org/wiki/MediaWiki.

Mukherjee R., Mao J., 2004, *Enterprise search: Tough stuff*, Queue, 2(2): 36.

O'Looney J., 2002, *Wiring Governments: Challenges and Possibilities for Public Managers*, Westport, Quorum Books, p.122.

Popadiuk S., Choo C.W., 2006, *Innovation and knowledge creation: how are these concepts related?*, "International Journal of Information Management", vol. 26, no. 4, pp. 302-312.

Reynolds G., 2010, *Information technology for managers*, Course Technology, Cengage Learning.

Richardson W., 2006, *Blogs*, *Wikis*, *podcasts*, *and other powerful Web tools for classrooms*, Thousand Oaks, CA Corwin Press.

Schneider D., DaCosta J., 2014, *Adding power to educational and research wikis with Semantic MediaWiki*, Retrieved 03.02.2014, http://edutechwiki.unige.ch/en/Adding_power_to_educational_and_research_wikis_with_Semantic_MediaWiki.

Semantic Mediawiki, *Properties and types*, Retrieved (January 2014) from http://semantic-mediawiki.org/wiki/Help:Properties_and_types.

Stewart T., Ruckdeschel C., 1998, *Intellectual capital: The new wealth of organizations*, "Performance Improvement", no. 37(7), pp. 56-59.

Tiwana A., 2000, *The knowledge management toolkit: practical techniques for building a knowledge management system*, Prentice Hall PTR.

Uden L., 2011, *Chapter 10 Semantics in Wiki*, [in:] Pardede E. (ed.), *Community-Built Databases Research and Development*, Springer-Verlag, Berlin–Heidelberg.

Weijermars R., 2011, *Building Corporate IQ – Moving the Energy Business from Smart to Genius*, Springer-Verlag, London.

SEMANTYZACJA WIEDZY INNOWACYJNEJ

Streszczenie: Artykuł porusza jeden z techniczno-organizacyjnych problemów tworzenia wiedzy dla innowacji, którym jest wspomaganie eksternalizacji i współdzielenia wiedzy w organizacji przy wykorzystaniu rozwiązań semantycznej sieci Web. Zaproponowano semantyczne podejście do gromadzenia i wykorzystywania wiedzy indywidualnej pracowników dla dobra organizacji. Podejście zaprezentowano na przykładzie zaczerpniętym z pewnej firmy branży metalowej. Przykład ma na celu wskazanie możliwości semantycznej integracji danych z różnych źródeł przy użyciu taksonomii przedsiębiorstwa w celu wspomagania współdzielenia wiedzy innowacyjnej przez pracowników.

Słowa kluczowe: semantic MediaWiki, taksonomia przedsiębiorstwa, wiedza innowacyjna, struktura semantyczna, semantyzacja.