

Social Ontology Documentation for Knowledge Externalization^{*}

Gonzalo A. Aranda-Corral¹, Joaquín Borrego-Díaz²,
and Antonio Jiménez-Mavillard²

¹ Universidad de Huelva. Department of Information Technology,
Crta. Palos de La Frontera s/n. 21819 Palos de La Frontera, Spain

² Universidad de Sevilla, Department of Computer Science and Artificial Intelligence.
Avda. Reina Mercedes s/n. 41012 Sevilla, Spain

Abstract. Knowledge externalization and organization is a major challenge that companies must face. Also, they have to ask whether is possible to enhance its management. Mechanical processing of information represents a chance to carry out these tasks, as well as to turn intangible knowledge assets into real assets. Machine-readable knowledge provides a basis to enhance knowledge management. A promising approach is the empowering of *Knowledge Externalization* by the community (users, employees). In this paper, a social semantic tool (called *OntoxicWiki*) for enhancing the quality of knowledge is presented.

1 Introduction

The competitiveness of companies active in areas with a high market change rate depends heavily on how they maintain and access their knowledge (i.e. their corporate memory) [5]. One of the main problems that current information systems have to face is that most of the information stored is not understood by machines, and so can't be effectively processed automatically. Semantic Web (SW) aims to extend the current Web by adding structured information. That information will be enriched with semantics, which will allow the description of the content on the Web, so that they can be interpreted by humans and machines. Ontologies are emerging as the most widely accepted technology for that purpose. Roughly speaking, SW's goal can be achieved by means of transforming information into Knowledge, using ontologies as a formal reference.

An ontology is a hierarchical collection of classes (concepts) and properties (relationships between these concepts) that models a particular conceptualization. For instance, the Business Enterprise Ontology¹ (EO) defines that "A Sale is an agreement between two Legal Entities for exchanging of a Product by a Sale Price. Commonly, the Product is a good or a service and the Sale Price is monetary, although another possibilities are included". Once formalized, these kind of definitions solve problems of interoperability at semantic level.

^{*} Partially supported by TIN2009-09492 project of Spanish Ministry of Science and Innovation, cofinanced with FEDER funds.

¹ <http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html>

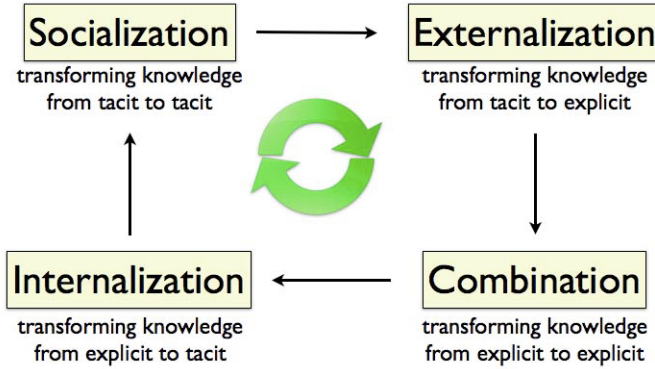


Fig. 1. Semantic Web Cake and Nonaka & Takeuchi's cycle

Whenever two information systems try to communicate and share information, semantic problems could arise (beside the obvious physical difficulties) and could be avoided. This is because these two systems “do not speak the same language”. Business management systems cannot avoid being affected by the consequences of this misunderstanding. Some enterprises use the term “Resource” while others talk about “Machine” to refer to the same concept. At other enterprises, however, “resources” as “raw materials” is used. If each enterprise defines or classifies their products in their own way, the result would be that automatic electronic transactions would be extremely difficult to achieve, and therefore e-business would not be possible. Ontologies provide common and shared knowledge from the described vocabulary through their classes and properties, in order to solve this problem. Its common language allows communication between individuals or information systems at different companies or organizations.

Nevertheless, the problem of semantic interoperability can persist on a different level, within the organization itself. It appears in the semantic heterogeneity among user's interpretations of ontology elements. Integration of information by means of ontologies is a current option at organizations which are aware of the knowledge value. There are powerful methods for integration [1] that produce robust ontologies, however, some problems related to *ontological literacy* of users can exist. Even though ontologies represent the knowledge of enterprises and users accept and consider them as their own, they may not understand that formal representation in a clear way (so they could underuse or misuse it). A proper solution is that in which own users can collaboratively describe the ontology usage and wiki technology is very appropriate for achieving it. A wiki is a software that enables the collaborative content creation. It allows users to create, modify or delete shared content quick and easily. Its purpose is to let multiple users editing web pages (called articles) related to a subject, so each one provides his/her knowledge and they can work together in order to complete the article. These users can create a community that shares content about common topics.

Since collective knowledge is extremely important for professional communities, a wiki has become as the essential tool that makes the collaboration among scientific communities, researchers, enterprises and any other user possible.

1.1 Motivation and Aims of the Paper

The use of novel methodologies for Knowledge management, as ontology-based ones, needs of a sound training of users (employees). A main semantic divide arises between ontology specification and a dialy use of ontologies at enterprises. The challenge to provide methods to solve this divide is the main motivation of this paper. Particularly, the aim is to present Ontoxicwiki, a tool designed for bridging this gap. This provides a social platform (based on wiki technologies) where users can document their usage of the ontology, always preserving its logical features and specification. It is importat to remark that OntoxicWiki is not a semantic wiki, it is a wiki for documenting ontologies instead. Ontoxicwiki provides a plugin for Protégé, an ontology editor that allows to document ontologies as well as to report ontology use cases. The user generated content is considered as the content of an ontology on documentation that the plugin merges with an ontology source. In this way nonspecialist users can generate information on the ontologies that their companies or social networks can exploit. These features distinguish Ontoxicwiki from semantic wikis, that are based on an underlying model of the knowledge described in its pages (while Ontoxicwiki is a wiki for documenting external knowledge represented by ontologies).

Structure of the paper. The next section describes how to combine Web 2.0 and SW to enhance Knowledge Assesment Management, addressing some of the requirements for SW tools. In Section 3 OntoxicWiki is presented. Main wiki features produced by OntoxicWiki are described in section 4. The paper ends with some conclusions and future work.

2 Knowledge Organizations and Semantic Web Tools

To create a machine-readable ontology, it is necessary to define it in a formal language (such as OWL²). As we move into such a technical and formal field, certain difficulties could arise for users. Some of them could be:

a) Although there are several ontology editors such as Protégé³, end users do not need to be skilled in these technologies. Every web user, both businessmen, scientists or researchers and users without a full knowledge should be able to manage these ontologies, especially, company employees. The syntactic and semantic complexity of ontologies can prevent its generalized usage.

b) Any user could design their own ontology, specifically suited for a given purpose, or reuse any available ones and take advantage of its previous usage. What ontology to choose? Where can he/she find testimonials of its usage?

² <http://www.w3.org/TR/owl2-overview/>

³ <http://protege.stanford.edu/>

c) There exist a number of undocumented ontologies, without any information about their application field or potential usage. Others, however, have been documented by their own creators (domain experts), which often make them difficult to understand and reuse, because of excessively technical jargon. This method of documentation is not usually appropriate for most users. It would be desirable to have mechanisms for documenting user experiences (for instance, cases of use, detected inconsistencies, etc.). This could be queried by future users and help them to select or understand the suitable ontology.

The above difficulties have an influence on any ontology-based process of externalization, organization and management of Knowledge. The primary source for building an ontology is a company's own information: databases, business process, etc. [1]. Therefore, it could be appropriate (as in [9]) to start analyzing from the perspective of Emergent Knowledge, roles and processes for Knowledge Asset Management (KAM) in creating knowledge organizations represented by Nonaka & Takeuchi's cycle [8] (see fig. 1). This cycle is based on four activities which transform the visibility, importance and value of KAM into organizations (*socialization, externalization, combination and internalization*). Knowledge Externalization is a key activity that create explicit intangible assets for the enterprises.

In SW, Knowledge is a tangible asset and the substance of processing. In Web 2.0 (W2.0), user generated knowledge is often based on the combination of different contributions from different users (or communities). Therefore, in Semantic Web 2.0 (SW2.0), similar KAM cycles could be studied in order to enhance them by means of new technologies or methods. In W2.0 networks, knowledge-creating communities are networks which are based on prosumers, it means, the creation and consumption of knowledge is entrusted to users. The Nonaka & Takeuchi cycle can be adapted to knowledge management in these networks, and some of these processes can be supported by Ontological Engineering theories and tools. Nonaka and Takeuchi's cycle projection in these networks shows four needs for creating truly SW2.0 communities: *emergent semantics, semantic user interfaces, knowledge networks and ontology alignment* (see fig. 1):

- The process of collaborative externalization by means of SW technologies is, in fact, a process of *emergent semantics* when tools for organizing knowledge are provided.
- Once Knowledge is externalized, in terms of ontologies, the combination of different knowledge sets is a problem of ontology alignment.
- Internalization of the Knowledge implies the fair use of SW tools which facilitate the creation -by employees- of common knowledge into their company, so semantic users interfaces are needed.
- Finally, the socialization of the knowledge produces knowledge networks when users adopt Web 2.0 behaviours.

Though these activities are similar to those used for constructing, mapping and managing collaborative knowledge spaces [6], innovative W2.0 tools and services will emerge.

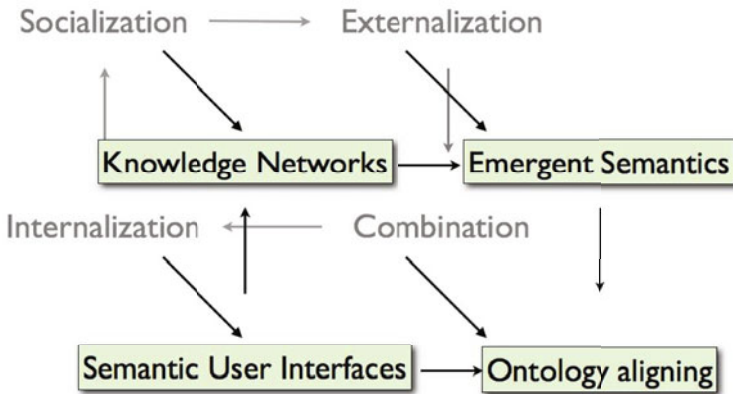


Fig. 2. Projection of the cycle

OntoxicWiki is designed to provide a semantic bridge between the knowledge activities of the cycle's projection, enhancing both Web2.0 and SW solutions in this context (fig. 3). Specifically, the tool is designed to satisfy several needs which arise from the problems described below. Firstly, Knowledge has to be represented in a comprehensive and friendly way, in order to be used by all type of users, no matter what their knowledge or skill are. Secondly, shared ontologies have to allow easy reuse, and finally, it must work on socially documented ontologies, which lets users read the contributions made by other users in order to facilitate their choice (and proper usage of) of a set of different ontology concepts, relations and individuals. These requirements lead us to propose a collaborative solution. It seems essential to have a web site where users could find shared ontologies with additional information. This allows us to learn interesting features about them, such as their scope, purpose or recommended applications. Thus, scientists, businessmen and other users will be able to access them, filtering the content by their respective areas of interest and, thanks to the added information, choose the ontology which best suits their work needs. OntoxicWiki supports these requirements: designing and documenting shared ontologies in a collaborative environment.

3 Ontoxicwiki

OntoxicWiki is a tool that has been born to bridge the gap between user and ontology. The main objective of this application is represent ontologies in an intuitive and easy understandable way for any user by providing with an environment from which repare and document ontologies socially, concretely a wiki. Supported by a Mediawiki platform, OntoxicWiki is a plug-in developed for the ontology editor Protégé. It has been designed this way due to Protégé can be easily extended by way of a plug-in architecture and a Java-based API for building knowledge-based applications with ontologies. Besides, it is a free, open-source

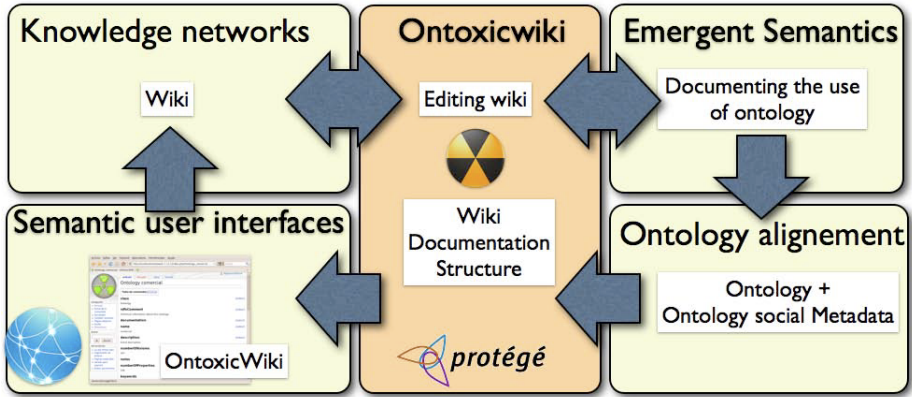


Fig. 3. Externalization of Knowledge by means of Ontoxicwiki

platform that provides a growing user community. In this sense, wiki has become the indispensable tool that best suits for collaboration among scientific communities, researchers, enterprises and other users.

3.1 Functionality

OntoxicWiki mainly consists of a toolbar and a web browser (see the technical architecture in Fig. 7). Its functionality is closely related to the lifecycle of ontologies in the context of this tool. In this scene, the lifecycle consists of five phases:

1. *Creating/loading the ontology in Protégé.* The process ends with an .owl extension file. This will be the target ontology. OntoxicWiki is not needed in this phase (Fig 6).
2. *Adding documentation to the ontology.* OSMV (Ontology Social Metadata Vocabulary) is inserted to the target ontology. Further details in section 4.
3. *Writing the ontology on the wiki automatically.* Information concerning to ontology resources, i.e. classes, properties and individuals, is extracted. Then, a wiki article is created for every class and property. The information will be written on this article depends on the type of resource is being considered, but each one will consist of a number of subsections, following the wiki-text syntax [<http://www.mediawiki.org/wiki/Help:Formatting>]. In case of a class, relevant data are its rdfs:comment, properties, subclasses, superclasses and instances. Each of these pieces of information is written on its respective subsection (rdfs:comment as natural language and the others as a list of elements). When describing a property, considered data are its rdfs:comment (in natural language), domain, range, superproperty and subproperties (lists of elements). Besides, special articles and subsections coming from OSMV

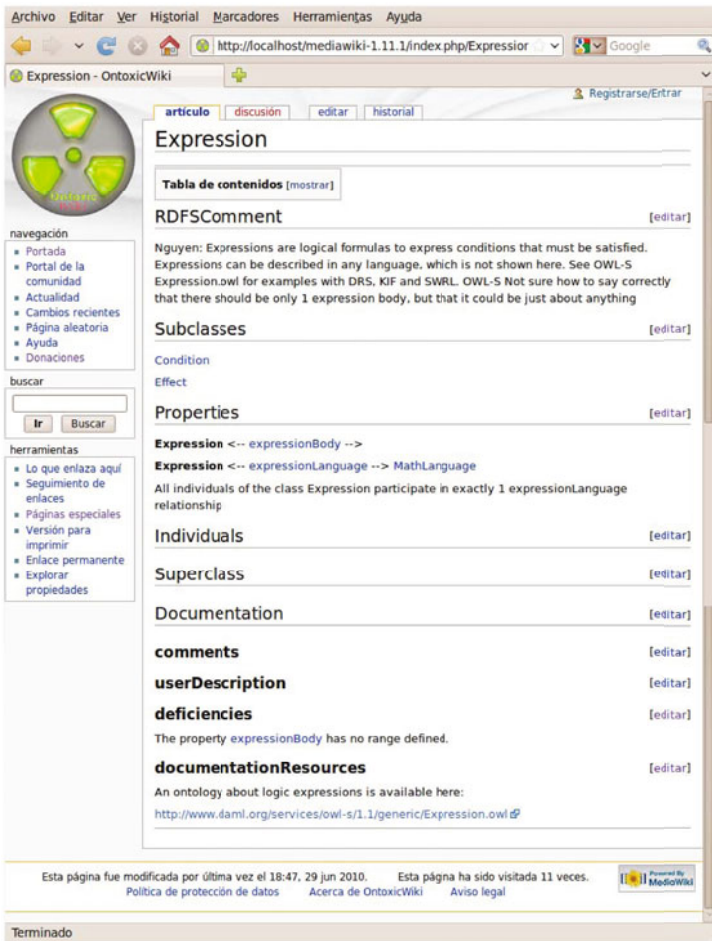


Fig. 4. Ontoxicwiki screenshot

are added to the original ones. Finally, articles are linked through hyperlinks to keep the original ontology structure (Fig. 5).

4. *Ontology revision/documentation*. Once ontologies are available on the wiki, the user only needs a web browser to access and navigate from one page to another. Users can visit an article (a class or property) just for reading or they can edit, delete it or create new ones (following some simple syntax rules for further processing). Revising and documenting is filling in the corresponding subsection. Like other wikis based on Mediawiki (e.g., Wikipedia), it is fully functional and allows, among others, actions such as registering new users or administration tasks. OntoxicWiki is not needed in this phase (Fig. 4).

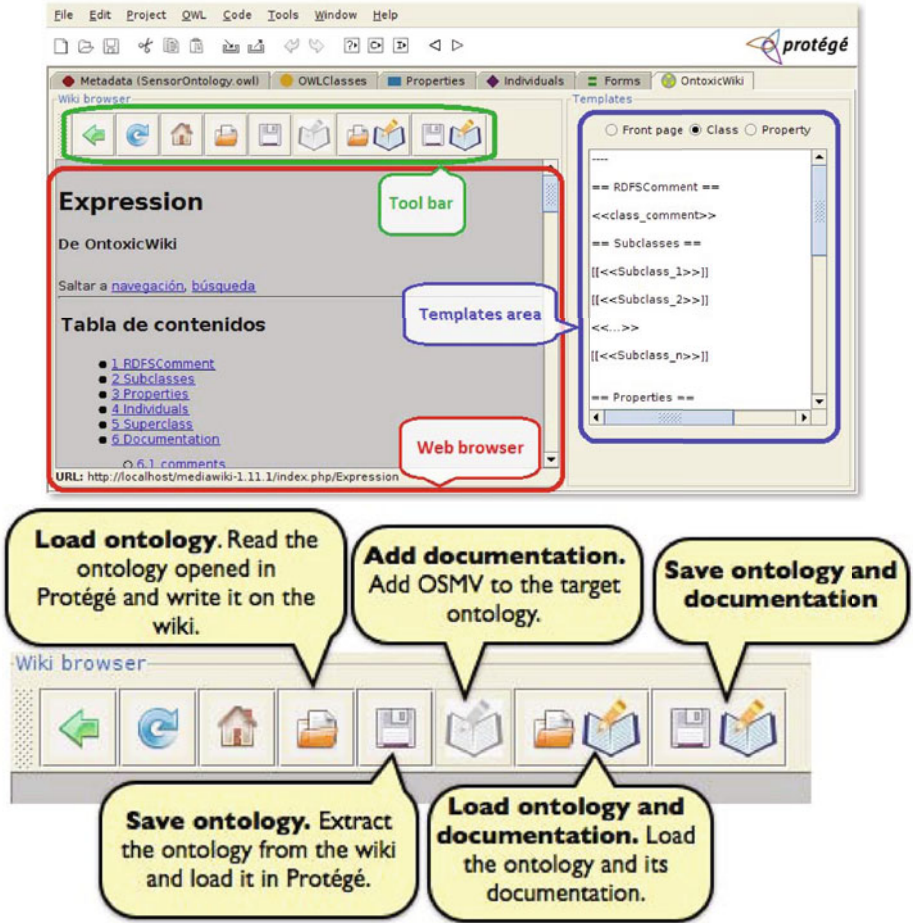


Fig. 5. Ontoxicwiki Protégé plugin and tool bar

5. *Retrieving the ontology from the wiki automatically.* Users can construct an OWL format ontology from wiki for later use. To achieve this, articles structure and relations (hiperlinks) between them are analyzed and classes and properties hierarchy is generated. Then the content of each article is read and parsed to build such class or property in the ontology. All information related to documentation (OSMV) is ignored, so only the target ontology is retrieved.

The user interface is mainly composed by three elements: a toolbar, the template area and the web browser. The toolbar (see Fig. 5, down) provides basic features for managing ontologies. Each button shows the related template. The template area displays the syntax corresponding to the wiki article.

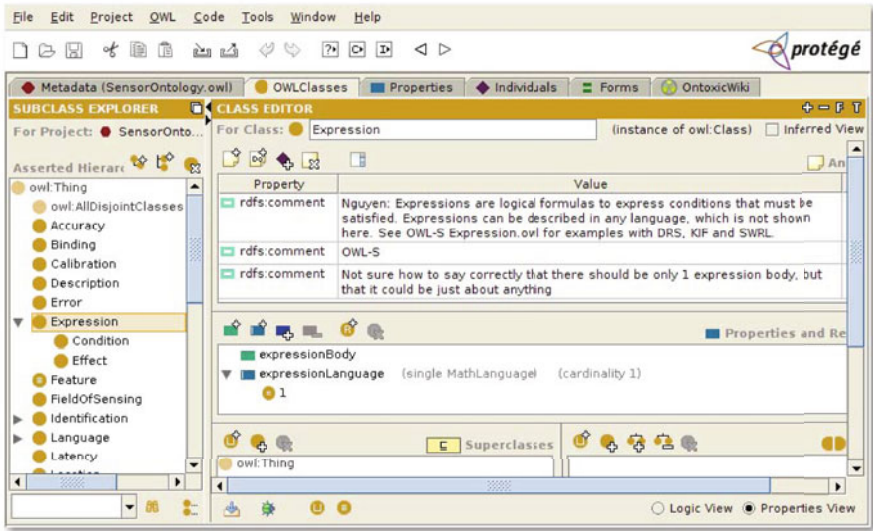


Fig. 6. Loading the ontology in Protégé

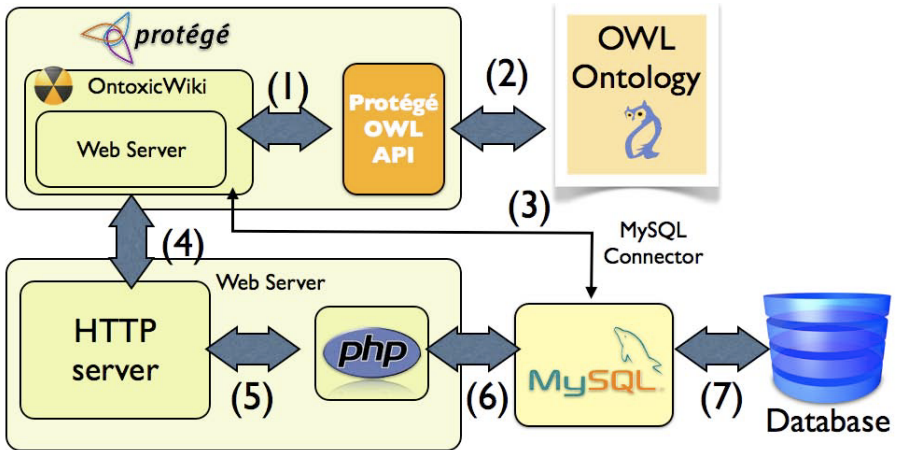


Fig. 7. Technical features of Ontoxicwiki (1) OW arch. (2) OW uses Protégé-OWL API to manage the ontology. (3) It accesses to database through MySQL Connector/J driver, which allows to connect from Java source to database. (4) OW browser makes requests to HTTP server. (5) HTTP server parses them and send to PHP engine. (6) It sends appropriate SQL statements to MySQL database manager. (7) Lastly, MySQL accesses database to read/write data and send back the results.

4 Features of the Wiki and the Integration of OSMV Ontology

An aim of Ontoxicwiki is to create a wiki which contains ontologies. It is hosted on a server with Mediawiki software installed. OntoxicWiki will connect to the database created by this software and will store the information related to ontologies on it. Once ontologies are available on the wiki, the user only needs a web browser to access and navigate from one page to another. Users can visit an article just for reading or they can edit it, delete it or create new ones (following the syntax rules for further processing). Like other wikis based on Mediawiki (for instance, Wikipedia), It is fully functional and allows, amongst others, actions such as registering new users or administration tasks.

Ontology Metadata Vocabulary OMV⁴ is a consensus ontology for Information Technology professionals, Computer Science and Knowledge Engineering. It consists of a set of instantiated classes and properties that provides useful data that describe ontologies. OntoxicWiki integrates an extension of OMV, called OSMV Ontology Social Metadata Vocabulary that contains classes and roles which specify social features. This extension allows users to provide two kinds of information. On the one hand, information about the ontology as a whole (domain, the number of classes, the language used to describe it, other similar ontologies on the Internet, etc). On the other hand, specific information for each class/property (user's comments or information sources for this class/property, etc.). In both cases, there exist data both technical and social nature. When OSMV is added to the target ontology, both ontologies are merged into one. The consequences are: OSMV's classes and properties are added to the target ontology's ones. An instance which represent the target ontology itself and an instance for each target ontology's class and property are created. As far as the wiki is concerned, new subsections are added to each article. On the front page, one subsection for each global documentation property appears: number of classes, language, similar ontologies, keywords, domain, author, etc. On each class and property articles, one subsection for each local documentation property is added: comments, user descriptions, deficiencies found and documentation resources for this topic, etc. From now on, OSMV provides slots to be filled in with global information relative to the ontology as a whole or to their classes and properties specifically.

5 Conclusions and Related Work

Ontoxicwiki has been designed for bridging the gap between the formal specification of ontologies and their users, by means of social documentation. There are similar approaches that combine Semantic Web and wiki technologies, although these have different scopes, applications or do not use the powerful features that ontology editors, such as Protégé, provide (for example, analysis of logical

⁴ <http://sourceforge.net/projects/omv2/>

consistency). As Ontoxicwiki were built as a Protege plugin, it has a valuable advantage over semantic wikis: the ontology consistency can be checked at any moment, just clicking on Protege's button. Obviously, an automated reasoner, i.e. Pellet⁵, must be installed and running. If Protege detects any inconsistency, OntoxicWiki could undo the changes based on the regular features of any wiki.

It is important to make clear that OntoxicWiki is not a semantic wiki. Semantic wikis provide the ability to capture or identify information about the data within pages, and the relationships between pages. Meanwhile, OntoxicWiki uses a wiki just to represent and document ontologies easy and intuitively. It is worth to remark that OntoxicWiki is useful for enhancing activities belong to Knowledge process cycle [10], because it provides a social platform where knowledge workers can capture their experience as well as they can access to documents through a nice GUI interface, namely wiki. Therefore, it is a promising research line to extend Ontoxicwiki to a complete suite for Semantic Knowledge Management that transforms ontologies in a clear way.

Semantic MediaWiki⁶ (SMW) helps to search, organise, tag, browse, evaluate, and share wiki's content. SMW adds semantic annotations that easily publish Semantic Web content. However, it is not designed for documenting ontologies, because it does not directly allow ontology debugging. Our tool is designed for a different purpose: Knowledge externalization of user experiences and social documentation of pre-existing ontologies (which were not built in a semantic wiki environment). For existing ontologies that were built using ontology editors, Ontoxicwiki helps to browse and to document them. From this point of view, its main feature would be to document and to report their use in companies.

OntoWiki⁷ allows users to navigate through ontology classes and properties like traditional wikis, but actually uses an underlying formally described ontology. Users can create instances of classes and relations and add their own documentation. It fosters social collaboration aspects by keeping track of changes, allowing users to comment and discuss every single part of a knowledge base, enabling to rate and measure the popularity of content and honoring the activity of users. However, it does not have a specific ontology for user documentation as OntoxicWiki.

AceWiki⁸ also uses a wiki to manage ontologies in an intuitively way and hides OWL and logic. In this case, the language is ACE, a rich subset of standard English that seems natural language but is in fact a formal language. But, unlike OntoxicWiki, AceWiki does not support user documentation.

A more specific approach is SWiM⁹, designed for editing/browsing mathematical ontologies [7]. One can easily identify the correspondences between the SW ontologies and mathematical ontologies.

⁵ <http://clarkparsia.com/pellet/>

⁶ <http://semantic-mediawiki.org>

⁷ <http://ontowiki.net>

⁸ <http://attempto.ifi.uzh.ch/acewiki/>

⁹ <http://kwarc.info/projects/swim>

It might be arguable that Ontoxicwiki has not user evaluation results in its first stage, since there is no difference from classic wikis. At design time, this was the critical feature. User evaluation enhances when Ontoxicwiki will be implemented as a Protege Server plugin and, thus, authors could use OntoxicWiki from any scientific community for collaborative edition and documentation of ontologies.

Lastly, it is likely that users will inconsistently document some classes/relations (even if ontology has been extracted from the company's own information). A solution is to use tagging, in order to apply semi-automatic reconciliation methods. For example, using agent-based methods [3]. Also, visual representations [4] could be considered, in order to provide an interface for ontology debugging, extending the features of OntoxicWiki for document debugging process.

References

1. Alexiev, V., Breu, M., Bruijn, J.d., Fensel, D., Lara, R., Lausen, H.: Information Integration with Ontologies: Experiences from an Industrial Showcase. John Wiley & Sons, Chichester (2005)
2. Aranda-Corral, G., Borrego-Díaz, J.: Ontological dimensions of Semantic Mobile Web 2.0. First principles. To appear in Handbook of Research on Mobility and Computing. IGI Press (2010)
3. Aranda-Corral, G., Borrego-Díaz, J.: Reconciling Knowledge in Social Tagging Web Services. In: Corchado, E., Graña Romay, M., Manhaes Savio, A. (eds.) Hybrid Artificial Intelligence Systems. LNCS, vol. 6077, pp. 383–390. Springer, Heidelberg (2010)
4. Borrego-Díaz, J., Chávez-González, A.M.: Visual Ontology Cleaning. Cognitive Principles and Applicability. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 317–331. Springer, Heidelberg (2006)
5. Fensel, D.: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, 2nd edn. Springer, Heidelberg (2004)
6. John, M., Melster, R.: Knowledge Networks – Managing Collaborative Knowledge Spaces. In: Melnik, G., Holz, H. (eds.) LSO 2004. LNCS, vol. 3096, pp. 165–171. Springer, Heidelberg (2004)
7. Lange, C., Kohlhase, M.: A Mathematical Approach to Ontology Authoring and Documentation. In: Carette, J., Dixon, L., Coen, C.S., Watt, S.M. (eds.) Calculemus 2009. LNCS (LNAI), vol. 5625, pp. 389–404. Springer, Heidelberg (2009)
8. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford Univ. Press, Oxford (1995)
9. Nykänen, O.: Semantic Web for Evolutionary Peer-to-Peer Knowledge Space. UPGRADE X(1), 33–40 (2009)
10. Staab, S., Studer, R., Schnurr, H., Sure, Y.: Knowledge Processes and Ontologies. IEEE Intelligent Systems 16(1), 26–34 (2001)