From data to knowledge through collaboration: bridging Wikis and knowledge systems

H. Astudillo[1], G. Canals[2], A. Diaz[3], A. Napoli[2], and M.G. Pimentel[4]

[1]Departamento de Informática, Universidad Técnica Federico Santa María (UTFSM)
Valparaiso, Chili (hernan@inf.utfsm.cl)

[2]LORIA (CNRS – INRIA – Nancy Université), Vandoeuvre-lès-Nancy, France
(Gerome.Canals@loria.fr,Amedeo.Napoli@loria.fr)

[3]LIFIA, Universidad de La Plata, Argentina
(alicia.diaz@lifia.info.unlp.edu.ar)

[4]Laboratório Intermidia Departamento de Ciências de Computação,
Universidade de São Paulo (São Carlos), Brazil (mgp@icmc.usp.br)

Abstract. *This paper presents some scientific aspects of the STIC AmSud project "Semantic-based support for Collaborative Design Activity". The main goal of this cooperation project is to define an operational environment for demonstrating how semantic technologies can be associated with collaborative design activities. In this way, semantic Wikis provide an original and operational infrastructure for efficiently combining both technologies.*

1. Introduction

This paper briefly presents the scientific aspects of a STIC AmSud project entitled as "Semantic-based support for Collaborative Design Activity" and involving researchers from Argentina, Brazil, Chili, and France. This project is aimed at integrating and sharing experiences and research efforts for defining an operational environment –including models, methods and tools– where semantic technologies are associated to collaborative computer-supported design activities. In particular, the project tries to demonstrate that semantic Web technologies are a suitable option to guide and improve a collaborative computer-assisted design process. Such an environment has to take advantage of domain ontologies for helping designers in assembling design components by guiding search for adequate documents, components, skills, and detection of conflicts as well. This research work involves competencies for: (i) knowledge engineering: knowledge representation, reasoning, knowledge discovery, ontologies and semantic Web technologies, (ii) support to collaborative design activities, with a focus on coordination and awareness, but also knowledge management for design process. The competencies of each team in the project are complementary w.r.t. the original topic of the project.

Researchers in the project met three times since the beginning of the project (April and October 2008, February 2009). For carrying on a proper research line in accordance with the original objective, it was decided to work on semantic Wikis with an application in the teaching domain, namely how to use semantic Wiki technology for designing a course on semantic Wikis themselves. For understanding this proposition, we introduce some background elements. A Wiki is roughly speaking a Web site (built around a set of pages) that can be edited by several people, possibly at the same time. Editing is

the collaborative task and problems of coherence and needs for synchronization appear while several people are working together. In this way, combining Wikis and semantic Web technologies is considered as a promising alternative for collaboratively creating and using information on the Web. The user-friendliness of Wikis as regarding multi-site content generation and the power of semantic technologies for organizing and retrieving knowledge may complement one another towards a new generation of Web-based content management systems. Accordingly, a semantic Wiki can be seen as a Wiki including an associated ontology, i.e. an operational representation model of domain knowledge, that can be used for annotating the content of Wiki pages and used for typing hyperlinks and testing consistency of contents (e.g. two users cannot state contradictory facts w.r.t. the underlying ontology). Moreover, an annotation process w.r.t. an ontology has a direct impact on knowledge access, semantic search and reuse, collaborative authoring, and social collaboration [Auer et al. 2006, Krötzsch et al. 2006].

In the present application context, i.e. designing a course on the topic of semantic Wikis, one has to be able to determine a set of resources –mainly textual documents but possibly extensible to multimedia support, e.g. video and sounds– to be analyzed and to set on a process able to extract useful units from the resources to be considered for designing the target course in being guided by a domain ontology. However, the transformation process from data to knowledge is hard to be fully automated: it may be oriented and guided by human collaboration at each step of the knowledge discovery process, i.e. when preparing data, tuning data mining algorithms, and interpreting extracted units. The process of turning the content of textual documents into knowledge units embedded in Wiki pages, and the different operations are illustrated on Figure 1. In this way, A parallel can be drawn with the way an analyst usually guides a knowledge discovery process [Lieber et al. 2008].

This is the purpose of this paper to introduce and explain a practical way of integrating semantic and collaborative technologies through a "semantic Wiki". Then, it is shown how to combine semantic techniques such as knowledge representation and knowledge discovery within a Wiki environment for guiding the transformation process from data to knowledge units for a practical application.

2. Knowledge discovery and representation: from resources to an ontology

Ontologies are the backbone of semantic Web in allowing software and human agents to communicate and to share domain knowledge [Antoniou and van Harmelen 2004, Staab and Studer 2004]. From a formal point of view, an ontology is considered as an explicit specification of a domain conceptualization. For being operational, an ontology has to be encoded within a knowledge representation language such as a description logic [Baader et al. 2003] or OWL [Antoniou and van Harmelen 2004]. From a practical point of view, besides ontologies, there exist different types of "ontological resources" such as thesaurus, vocabularies, dictionaries, collections of documents, and databases. Every ontological resource provides a specific aspect of domain knowledge. For taking into account these ontological resources, a framework can be designed in which the content of resources can be integrated for being used as a "domain knowledge container" for knowledge sharing and reuse.

Following this way, Formal Concept Analysis (FCA) and its extension Rela-

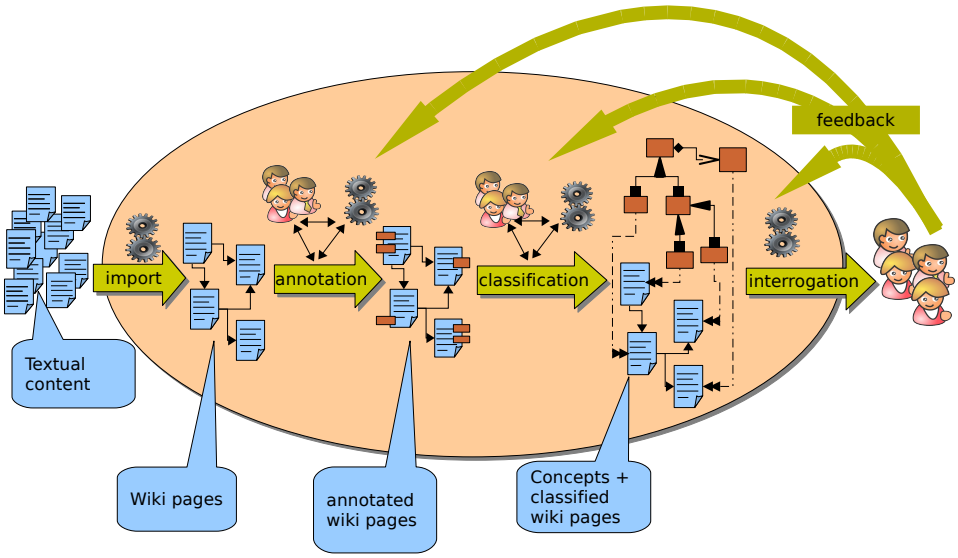

Figure 1. The transformation process from textual content to knowledge units.

tional Concept Analysis (RCA) can be used for designing and completing an ontology from a set of resources (especially textual documents as in [Bendaoud et al. 2008b, Bendaoud et al. 2008a]). FCA and RCA lead the transformation process between ontological resources and a concept lattice supporting an "ontological schema", i.e. a set of related concept definitions. One way of processing is to build a "source ontology" from resources and then to extend this source ontology by progressively adding new units extracted from additional resources. The transformation process can be based on FCA operations (e.g. apposition) and on RCA operations (e.g. extracting relations between concepts). Finally, elements in the resulting concept lattice can be represented within a knowledge representation language such as OWL for obtaining a concept hierarchy representing the ontological schema given by the concept lattice. In this way, FCA and RCA are "core processes" in the design of a target ontology from a set of heterogeneous resources. Firstly, FCA and RCA take into account all basic elements within an ontology, such as objects (or individuals), attributes, and relations, i.e. roles with a domain, a range, and possibly a cardinality and a quantification. Secondly, FCA provides operations for creating, managing, and updating concept lattices. When the resulting concept lattices are transformed into concept hierarchies, a classifier can be used for classification-based reasoning, content-based information retrieval, and answering queries.

This is an operational way of designing ontologies from a set of resources but this process is hard to be fully automated and requires a human intervention. This calls for collaborative architecture such as provided by a Wiki within which the transformation process from resources to knowledge units is explained in the next section.

3. Wikis and Semantic Wikis

A Wiki is a collection of Web pages allowing collaboration and user-contributed "knowledge production", by enabling users to contribute or modify content of pages

[Cunningham and Leuf 2001]. One of the best-known repository of user-contributed knowledge is Wikipedia, also one of the largest and fastest growing online sources of encyclopedic knowledge [Krötzsch et al. 2007]. The richness of its embedded structural information is mainly based on hyperlinks, with relations such as synonymy, polysemy, and additional tools such as infoboxes and templates. A Wiki offers simplicity and a social dimension, and emergence of structured knowledge repositories of collaborative nature.

Here, knowledge has to be understood as "knowledge for human agents" and not for software agents. This is why, for guiding a coherent development of a Wiki infrastructure, semantic technologies and especially ontologies can be used within Wikis, leading to semantic Wikis [Auer et al. 2006, Krötzsch et al. 2006, Buffa et al. 2008]. Semantic Wikis allow knowledge processing for humans and machines. The Wiki infrastructure can be used for supporting an ontology: the Wiki can be considered as a front-end of the ontology maintenance system, with Wiki pages as concepts, typed links as relations and attributes. By contrast, ontologies can be used within Wikis for supporting page selection, annotation (tag organization), searching, querying, faceted navigation, guided edition, and consistency checking. For example, the semantic mediawiki system addresses consistency of contents, knowledge access, and reusing information [Krötzsch et al. 2006]. Moreover, an annotation process w.r.t. an ontology has a direct impact on querying information: annotations can be categories, relations, and attributes, and can be represented as logical statements manipulated within reasoning schemes. In the same way, three main tasks linked to Wikis are improved by the presence of an ontology, namely collaborative authoring (editing), social collaboration (change tracking), and semantic search (browsing) [Auer et al. 2006].

4. Combining knowledge discovery, ontologies, annotation, and semantic Wikis

There are two main views relating Wikis and ontologies: (i) Wikis for improving ontology infrastructure, and (ii) ontologies for improving the development and the management of Wikis.

- In the first view, collaboration plays an important role for editing pages and for gathering and integrating resources of different types. Knowledge discovery (KD) processes can be applied to such a container of resources for extracting units. These units, after interpretation by an analyst, can be embedded within an ontology. In this view, a semantic Wiki can be used for selecting, collecting, and preparing data (documents) in a collaborative way for ontology design and extension.
- In the second view, an ontology plays the role of a domain model, providing a "domain terminology" with terms and associated meaning. In this view, improvements for the Wiki activity are mainly based on document searching –searching by content– and understanding, coherence checking, and guided editing. Both views involve collaborative aspects and "knowledge production" (i.e. extraction or creation): collaboration based on a Wiki guides knowledge organization and evolution in the first view while an ontology can be used for controlling the evolution and checking the consistency of the new elements brought through collaboration in the second view.

Actually, a semantic Wiki can be used as a support for knowledge discovery and knowledge management by combining two interrelated views of Wikis and ontologies. The following loop summarizes the operations that are under study within the research group and that will be tested on the design of a course :

- Based on an initial ontology, a set of documents of interest "ranked" by their content w.r.t. a given topic is selected.
- Then, documents are annotated and related (through hyperlinks) in a collaborative way and under the control of the ontology, i.e. using terms defined within the ontology.
- The resulting set of annotated documents can be analyzed using data mining algorithms for extracting elements of interest.
- These new elements can be interpreted and then proposed as knowledge units for extending and improving the initial ontology. The Wiki infrastructure can be used for making easier interactions.
- The loop is closed: starting from an ontology and going back to the ontology through a collaborative activity within a semantic Wiki using knowledge discovery and knowledge representation techniques.

This process can be seen as a slight adaptation of the transformation process given on Figure 1. An example on the design of a course will allow us to instantiate this design loop. The following operations can be carried out and the content of the course could be roughly the following:

- Presentation of a running example and its context: e.g. organization of the collections in a library or in a museum and visits.
- What is a Wiki and what are the principles of "computer supported cooperative work" (CSCW).
- What are semantic technologies, knowledge representation, reasoning, ontologies, annotation, and semantic Web.
- How Wikis meet semantic technologies, or alternatively how CSCW meet semantic Web: anatomy of a semantic Wiki (e.g. reasoning tools, storage, querying).
- Case study: collaborative course design.
- Advanced topics: e.g. knowledge discovery guided by domain knowledge, good practices in practical aspects, study and use of feedback and of recommendations.
- Conclusion, generalization, and reuse aspects.

5. Conclusion

In this paper we presented some scientific aspects of a STIC AmSud project entitled as "Semantic-based support for Collaborative Design Activity". One objective of this research project is to define an environment where collaborative and semantic technologies meet for improving design activities. We are currently building a framework for building new courses from a set of resources (such as pedagogical documents) using collaboration, knowledge discovery, and annotations processes, supported by the infrastructure of a semantic Wiki. The semantic Wiki infrastructure allows guided interactions between users, documents, and ontologies. The underlying knowledge system, combining knowledge discovery and knowledge representation processes, guides and improves information retrieval, resource classification and annotation, course design, and reasoning for problem solving and evaluation.

At the moment, people in the research project are working on the details of the design process both from theoretical and practical points of views. One important thing is to deeply understand the interactions of semantic and Wiki technologies, and to formally describe the role from each sides. Embedding collaborative aspects of Wiki and semantic techniques in association with knowledge discovery and representation (text mining, annotation) determines a new research direction that has still to be fully investigated.

The existence of this AmSud research project gives us the opportunity of working together with different and complementary backgrounds on these aspects with the hope of obtaining substantial results in a next future.

References

Antoniou, G. and van Harmelen, F. (2004). *A Semantic Web Primer*. The MIT Press, Cambridge, Massachusetts.

Auer, S., Dietzold, S., and Riechert, T. (2006). Ontowiki - a tool for social, semantic collaboration. In *The Semantic Web – ISWC 2006*, LNCS 4273, Springer, pages 736–749.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., and Patel-Schneider, P., editors (2003). *The Description Logic Handbook*. Cambridge University Press, Cambridge, UK.

Bendaoud, R., Napoli, A., and Toussaint, Y. (2008a). Formal concept analysis: A unified framework for building and refining ontologies. *Proceedings of the 16th EKAW*, LNCS 5268, Springer, pages 156–171.

Bendaoud, R., Toussaint, Y., and Napoli, A. (2008b). Pactole: A methodology and a system for semi-automatically enriching an ontology from a collection of texts. *Proceedings of ICCS 2008*, LNCS 5113, Springer, pages 203–216.

Buffa, M., Gandon, F. L., Ereteo, G., Sander, P., and Faron, C. (2008). Sweetwiki: A semantic wiki. *Journal of Web Semantics*, 6(1):84–97.

Cunningham, W. and Leuf, B., editors (2001). *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley.

Krötzsch, M., Vrandecic, D., and Völkel, M. (2006). Semantic mediawiki. In *The Semantic Web – ISWC 2006*, Lecture Notes in Computer Science 4273, pages 935–942. Springer.

Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., and Studer, R. (2007). Semantic wikipedia. *Journal of Web Semantics*, 5(4):251–261.

Lieber, J., Napoli, A., Szathmary, L., and Toussaint, Y. (2008). First elements on knowledge discovery guided by domain knowledge (KDDK). *Concept Lattices and Their Applications*, LNAI 4923, Springer, pages 22–41.

Staab, S. and Studer, R., editors (2004). *Handbook on Ontologies*. Springer, Berlin.