Collective Cognition with Semantic Mediawiki: Lessons and Experiences

(Extended Abstract)

Jie Bao, Li Ding, James Hendler
Tetherless World Constellation, Department of Computer Science
Rensselaer Polytechnic Institute, Troy, NY, USA
{baojie,dingl,hendler}@cs.rpi.edu

Introduction

Web-based tools have fundamentally revolutionized the ways of knowledge acquisition, propagation, aggregation, understanding, and analysis. This is exemplified by Wikipedia, one of the most successful Web tools that support collective knowledge management in a large user community. Semantic wikis are extensions to wikis with semantic technologies that are aimed at further facilitating better human understanding and automated knowledge processing. Common features of semantic wikis include the incorporation of semantic markup into wiki pages which can be further translated into Semantic Web languages, e.g., RDF or OWL, and querying mechanism for better knowledge retrieval and propagation. Therefore, semantic wikis are capable of supporting two kinds of knowledge representations: informal knowledge represented as natural language text (as normal wikis support) for human reading, and formal knowledge representation with explicitly defined semantics that is primarily for automated machine processing. Among all semantic wiki systems, Semantic Mediawiki (SMW) [2] is probably the most successful one. We therefore use SMW as the representative of semantic wikis for our study and evaluation.

To reproduce the success of Wikipedia, a semantic wiki needs to follow some proven working principles, such as collaboration and ease of use. However, our experiences with several real-world projects reveal that semantic modeling is substantially more challenging than writing a conventional wiki for common users. The goal of this paper is to summarize some of our observations on key issues in collective cognition using SMW, including the choice of knowledge modeling patterns, the context and organization of knowledge, and collaboration protocols.

Knowledge Modeling

The native formal knowledge representation in SMW is in forms of categories and typed links (a.k.a. properties). They roughly correspond to RDF triples. For instance, on a page "x", the following scripts have RDF correspondences:

(category)	[[Category:Person]]	=>	x	rdf:type	Person.
(typed link)	[[knows::JaneSmith]]	=>	x	knows	JaneSmith.

In the deployment of SMW, it is often observed that there is only a small group of users that can think semantically and create knowledge models in the RDF-style. On the other hand, the majority of users on a wiki, even after some training, are often confused by the RDF-style modeling (e.g., the difference between categories and typed links, and the choice of vocabulary), and tend to less actively involved with direct annotation using the SMW scripts after some failed attempts. This leads to the choice between two common knowledge models on a semantic wiki, i.e., arbitrary RDF-style semantic markup that is heavily used by a selected few elite group, and "schema"-based modeling, often represented in the form of pre-defined wiki templates, that are used by "common" users of the wiki to access data via forms or prebuilt queries.

Some of the best practices of SMW are that of the combination of the two types of models. The stable and commonly agreed knowledge structure (e.g., basic information about a person) would be represented as a template-based schema, while users with extra needs for semantic modeling may choose to add arbitrary RDF-style markup.

In the course of growing of a wiki, the schema may evolve while some of the arbitrary RDF-style annotations become commonly useful. In this regard, semantic wikis are often similar to evolving collaborative databases.

We may regard this evolving process as a collective cognition process where the "schemas" represent stable, shared knowledge, and the arbitrary RDF-style markup (often inline with natural language texts) represents less structured and less shared knowledge. While end users will not typically edit with the RDF-style markup, both results of the two types of knowledge models will be shown in the UI, e.g., fact box and semantic browsing. For end users, the strength of a semantic wiki is often not in that it can better *capture* knowledge than normal wikis do, but in that the semantic annotations help in *abductive understanding* of the knowledge domain.

A large semantic wiki requires, like a database-based software project, careful design and documentation of the "schema" models, typical workflows and frequently used queries. In addition, it also requires knowledge engineering efforts and appropriate documentation, e.g., about the core category and property hierarchy. Those documents are among the key enablers for the success of a SMW-based project.

Knowledge Organization and Context

While there has been some work on context models of other semantic wikis [1], little attention has been paid on the context model of SMW. A context model determines where to write and where to find knowledge on SMW, and preserve the provenance information about knowledge assertions. The lack of a context model is observed with some problems in practice.

SMW (and many other semantic wikis) organizes knowledge into *pages*, this poses two major differences from RDF: 1) in RDF a statement can be asserted at everywhere, but in SMW it can only be put on a triple's subject page; 2) an RDF statement can be about arbitrary URIs, but in SMW a triple's subject and predicate can only be a wiki page. The page-centric organization has been found unsatisfactory sometimes:

- It may require the creation of many trivial, small pages. During the editing process, users often feel tedious to navigate through a couple of pages even for a simple task.
- It is troublesome to describe things (e.g., an external URL) that have no corresponding wiki pages.
- Due to the difficulty of determining where to write knowledge (i.e., the best "subject" pages), many users tend to use only the wiki editing functionalities without semantic annotations, often resulting in long pages containing many facts written in natural languages.
- Many users are confused of query-based pages: they do not know how to track the source of the queried results when they want to change a query-based page.

We argue that this would be remedied with introducing a context model to SMW, such that a triple can be asserted more freely, and its provenance can be tracked in editing and displaying. Besides, there is a need for representing an entity without explicitly giving it a page name, similar to blank nodes in RDF.

Collaboration Protocols

Reaching a consensus of knowledge cognition on a semantic wiki relies on a well designed collaboration protocol. On Wikipedia, one of the keys for establishing commonly acceptable description is the NPOV (Neutral Point of View) principle, which allows *multiple* points of view co-exist on one page with verifiable, authoritative *sources*.

Due to the inherent difference among individual cognition processes, knowledge statements from different users are often conflicting to each other. However, currently there is no built-in mechanism in SMW to support a collaboration protocol like NPOV: a triple is either true or false, and there is no way to formally describe the source of a triple. Unlike in Wikipedia where natural language text can accommodate and explain multiple points of view on a single page, the formal knowledge model of SMW lacks the support for such reconciliations. This has led to an

interesting new form of wiki "edit war": the ontology war, i.e., different users may try to impose their views on the ontological model (e.g., category hierarchy) by repeatedly reverting each other's contributions. Ontology warring is extremely dangerous when it involves terms that are used in popular templates or queries.

In practice, ontology wars may be restricted by separating an ontology into smaller, more personalized ontologies and allowing mapping between them. This has been practiced on the Tetherless World Wiki[1]. We believe a more principled solution may follow one of the following approaches:

- To have categories and typed links optional contextualized by authors, similar to the tag contextualizing mechanism in delicious and flickr.
- As we mentioned earlier, to introduce a provenance model of SMW knowledge statements, so that different versions of truth may be formally represented with explicitly given sources.

Conclusions

In this paper we analyzed some key aspects of using Semantic MediaWiki for collective cognition, including knowledge modeling, organization and collaboration protocols, drawing our lessons and experiences from several real-world projects using SMW. Our key observations are:

- Modeling in SMW can be regarded as an evolving cognition process where a knowledge schema, represented as templates, is often needed for end users to easily interact with the system.
- A successful semantic wiki is mixed efforts of both software engineering and knowledge engineering, both requiring some documentation.
- We observed several differences between knowledge management in SMW and RDF-based knowledge management in general.
- We showed that a context/provenance model is in need for SMW to support better knowledge organization and collaboration protocol. This will be our main future work.

References

[1] ElstLudger van Elst, Malte Kiesel, Sven Schwarz, Georg Buscher, Andreas Lauer, Andreas Dengel. Contextualized Knowledge Acquisition in a Personal Semantic Wiki. In Proceedings of the 16th international Conference on Knowledge Engineering: Practice and Patterns, 172-187. 2008.

[2] Markus Krötzsch, Denny Vrandecic, Max Völkel, Heiko Haller, Rudi Studer: Semantic Wikipedia. Journal of Web Semantics 5/2007, December 2007.

[1] http://tw.rpi.edu/wiki/Category:Public_Ontology