SOBOLEO – Social Bookmarking and Lighweight Engineering of Ontologies

Valentin Zacharias
Forschungszentrum Informatik, FZI
Haid-und-Neu Strasse 10-14
76131 Karlsruhe, Germany

zacharias@fzi.de

Simone Braun
Forschungszentrum Informatik, FZI
Haid-und-Neu Strasse 10-14
76131 Karlsruhe, Germany

braun@fzi.de

ABSTRACT

We present SOBOLEO, a system for the webbased collaborative engineering of SKOS ontologies and annotation of web resources. SOBOLEO enables the simple creation, extension and maintenance of taxonomies. At the same time, it supports the annotation of web resources with concepts from this taxonomy.

Categories and Subject Descriptors

H.5.3 [**Information Systems**]: Group and Organization Interfaces; H.3.3 [**Information Systems**]: Information Search and Retrieval; H.3.5 [**Information Systems**]: Online Information Services; H.4.3 [**Information System Applications**]: Communications Applications

General Terms

Design, Human Factors

Keywords

Ontology engineering, social software, collaboration, annotation, SOBOLEO

1. INTRODUCTION

So far, there is no integrated tool, which supports the collaborative creation of an index of relevant web resources together with the shared vocabulary used to organize it. Today's tools either allow the creation of a shared vocabulary or to structure information resources according to some fixed vocabulary. In addition, most tools for editing a shared vocabulary are tailored to people with training in knowledge engineering and not for domain experts. That results in a number of problems for using such tools:

1. Involving engineering experts for knowledge formulation and modelling causes the creation of a shared vocabulary to be more expensive [1].

2. Knowledge engineers often have only a limited understanding of the domain and thereby introduce more errors [1].

3. Organizational barriers arise from the fact that the group of people creating a vocabulary is different from the group using it. These barriers cause communication problems: new requirements are collected from the users and need to be understood by the knowledge engineers; in turn, changes to the shared vocabulary need to be explained to the users [9].

4. The use of different sets of tools for the creation/maintenance and for the use of a structured vocabulary often results in less frequent updates to the shared vocabulary. This again causes time lags between the moment a new concept emerges in the domain and its inclusion in the vocabulary. It has been shown that even modest time lags can make structured vocabularies unusable for fast moving domains [9].

For these reasons, tools for collaborative creation of shared vocabularies need to be usable by the people in the domain themselves. And—at least in fast changing domains—there must be a simple, quick and seamless way to change the shared vocabulary underlying an application.

2. THE SOBOLEO SYSTEM

In order to satisfy these requirement we developed SOBOLEO. SOBOLEO is the acronym for **So**cial **Bo**okmarking and **L**ightweight **E**ngineering of **O**ntologies. The system's goal is to support people working in one domain in the collaborative development of a shared vocabulary and of a shared index of relevant web resources. With SOBOLEO it is possible to create, extend and maintain taxonomies according to the SKOS Core Vocabulary [2] in a simple way. It also supports the collection and sharing of relevant web resources (bookmarks). These bookmarks can be annotated with concepts from the SKOS taxonomy or arbitary tags for better retrieval. One instance or installation of SOBOLEO is meant to be used by a community with interest in building a shared vocabulary and a web index. Within this one instance, the users create and maintain one taxonomy and one shared index of web resources collaboratively.

3. SYSTEM FUNCTIONALITIES

SOBOLEO is based on AJAX technology and works in most current browsers (we tested Internet Explorer 6 and 7, Opera 9, Firefox 1.5 and 2.0). It consists of four major parts: (1) a collaborative real time editor for changing the taxonomy, (2) a tool for the annotation of web resources, (3) a semantic search engine for the annotated web resources, and (4) a taxonomy browser for navigating the taxonomy and the content of the web resources index. Editing the taxonomy
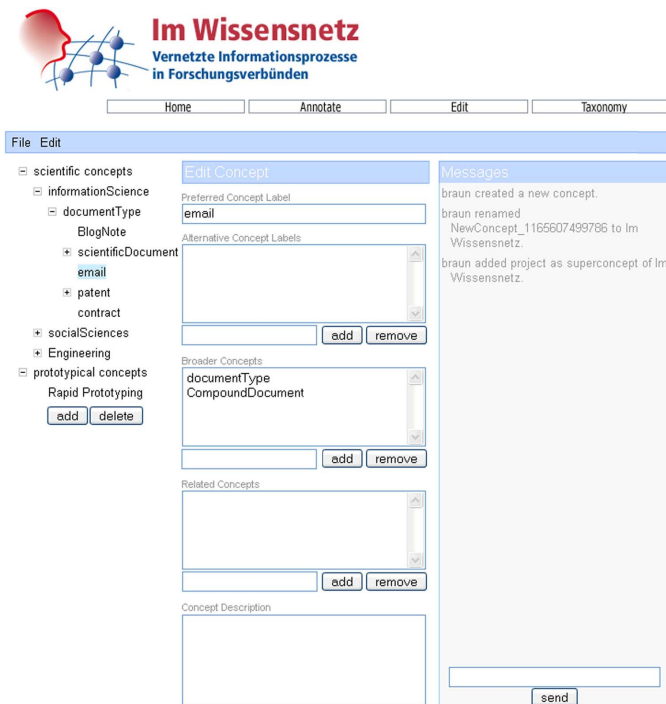
Figure 1: *The editor for the structured vocabulary.*

Figure 2: *The annotation pop up.*

or the index of web resources requires authentification. The other areas are open accessible. In the following paragraphs we will describe each of these parts.

3.1 Editor

SOBOLEO uses a subset of SKOS [2] as the format for the structured vocabulary or taxonomy. It supports concepts and broader, narrower and related links between them. Concepts have a preferred label and a description; they can have any number of alternative labels.

Entering the editor after succesfull authentification, a user sees the current state of the taxonomy and the changes of other users in realtime.

The interface of the editor (see Figure 1) consists of three main parts. To the left of the screen is a tree display of the taxonomy. It shows the preferred labels of the concepts and the tree formed by the narrower-broader-relations between them. Users can drag'n'drop concepts on other concepts to create new broader relations or drag'n'drop while pressing shift to remove them[1]. The operations cut, copy and paste (and their shortcuts) are also supported on the concepts in the taxonomy. Buttons below the taxonomy tree allow for the creation and deletion of concepts.

Clicking on a concept displays its details in the center part of the screen. The input fields at the center of the interface can be used to edit the relations between concepts, the labels and the description of the concept. The editing of relations between concepts is supported by auto completion of entities in the taxonomy.

The right part of the screen is reserved for message. The message pane displays automatically generated messages for all changes made to the taxonomy by any user. At the same

time, the message pane acts as chat window and shows all exchanged chat messages. An input field below allows the user to send chat messages to other users currently editing the taxonomy.

3.2 Annotation of Web Resources

Clicking on the annotate button anywhere in SOBOLEO opens the annotation interface (see Figure 2) as a pop up. In this interface, users can enter the url and the title of the document they want to add to the index. In another input field they can add a number of concepts from the taxomomy or arbitary tags to annotate this new resource. This topic field offers auto completion of entities in the taxonomy[2] to encourage users to apply these concepts. Tags not previously known to the system can be used as well and will appear in the taxonomy editor as "prototypical concepts"; users can consolidate and move them within the taxonomy later. In this way we allow for the seamless gathering of new concept ideas when they are occuring.

A better way to use the annotation tool is, however, to drag the annotate button on to the browser bookmarks, where it will work as a bookmarklet. Whenever a user visits a site she deems interesting for the index, she can click on the bookmarklet and it will open the annotation interface with url and title already filled in based on the page that is open in the browser.

The SOBOLEO system automatically sends out a crawler to any newly added url to fetch and index the contents of the page[3]. The contents of the page together with any concepts used to annotate it gets added to the index.

Whenever the annotate interface is opened for a particular url, the system checks whether there is already an annotation. If so, the annotation is displayed and can also be

---

[1]Removing broader relations in this way is currently not working in Firefox

[2]Based on the preferred label of the concepts

[3]Please note that SOBOLEO cannot add the content of sites that use frames or require authentification—in these cases only the title, the url and the annotated concepts will be saved

Figure 3: *The search result page.*

edited. In the current system one instance of SOBOLEO has one annotation per url. This annotation consists of a title and an arbitrary numbers of concepts. Annotations are maintained collaboratively.

3.3 Searching

SOBOLEO includes a semantic search engine supporting search and retrieval of the indexed web resources. The SOBOLEO search engine is used by typing search terms into a text field—similar to common internet search engines. The entered search string, however, is analyzed for occurrences of concept labels from the taxonomy. An occurrance of a label of a concept in a search string is understood as a *reference* to a concept. If references to concepts can be found, the SOBOLEO search engine searches for web resources annotated with these concepts or narrower ones (related concepts are currently not regarded). The results from the semantic search are combined with the result from a full text search in the contents of all annotated pages.

On the result page (see Figure 3) the SOBOLEO search engine gives feedback on which concepts it understood the query to be referencing. Depending on the search string, the results and the taxonomy, the system may also propose a number of query refinements or relaxations. The interface lists all found resources with their title linking to the original page, with annotated concepts, a short excerpt of the page content highlighting the search terms and the exact url. The edit links for each result allow to quickly modify or remove the annotation of a web resource. The annotation pop up interface is used (see Figure 2).

3.4 Taxonomy Browsing

With the browsing interface users can navigate through the taxonomy and the content of the web resources index

Figure 4: *The taxonomy and associated web resources browser.*

(see Figure 4). Starting from the root concepts, the users can click through the taxonomy concepts. On top, the users see the currently selected concept with its preferred and alternative labels and its description. Additionally, all its broader, narrower and related concepts are displayed as links for further navigation. Underneath the concept details there is a list of all resources which are annotated with the currently selected concept or with one of its narrower concepts. These resources are further ranked by their date they were collected, thus the newest resources appear upmost.

4. IMPLEMENTATION

The server side of SOBOLEO is implemted in Java on top of the Apache Tomcat [16] application server. The system uses Jena [3] for RDF processing, Hibernate [10] and HSQLDB [12] for persistance. The index of web resources and the semantic search is build on top of the open source framework Lucene [14]. The client side AJAX interfaces are created with Google's GWT framework [8].

5. RELATED WORK

A large number of other approaches for the creation of structured vocabulary and the building of a shared web index exist. We consider the semantic wikis (i.e. [13, 15]) and social web annotation tool like del.icio.us to be the most relevant (i.e. [6, 11, 5, 7, 4]).

Compared to Semantic Wikis it is immediately obvious that SOBOLEO is much more spezialized. Semantic Wikis are general purpose knowledge formulation tools—SOBOLEO is deliberately restricted to a simple subset of SKOS.

Because of this restriction SOBOLEO can offer more specialized user interfaces and search facilities making it faster, simpler and more effective for the creation of a shared web index together with the taxonomy used to organize it. SOBOLEO's editor also allows for a closer collaboration among the editors and virtually eliminates the chance for edit collisions. On the other hand, however, it is unclear whether SOBOLEO's interface could scale to knowledge models of a size that rivals Wikipedia.

Compared to state-of-the-art social web annotation tools like del.icio.us, SOBOLEO offers more ways to structure the shared vocabulary that is used for the annotation. The state-of-the-art structure in such tools is still formed by unconnected tags. In some cases (i.e. [6], [11]) it is possible to build up a structure by grouping tags underneath a super-tag. However, this functionality is very restricted and structures like super-tags of super-tags are not further regarded. Moreover, such a structure is built up on top—it is private and not shared. Another major difference is that tools like del.icio.us are geared to more diverse communities; it simply would not make sense to try to develop one shared vocabulary for all web users. Whereas SOBOLEO is meant for smaller, integrated groups working in one field.

6. CONCLUSIONS AND FUTURE WORK

We have presented the SOBOLEO system that makes it easy with its completely web based interface to jointly create a structured vocabulary and an index of web resources that is organized with this vocabulary. It combines a collaborative real time editor for a subset of SKOS with an easy to use annotation interface and a semantic search in the annotated resources.

We believe that SOBOLEO is already a useful tool, however, it is not as powerfull as we would like. For the near future we plan:

- Better evolution support for the editor, including undo and joining of concepts.

- Better support for the social aspects of collaboration, like private annotations or multiple communities within one SOBOLEO instance.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] K. Barker, V. K. Chaudhri, S. Y. Chaw, P. Clark, J. Fan, D. Israel, S. Mishra, B. W. Porter, P. Romero, D. Tecuci, and P. Z. Yeh. A Question-Answering System for AP Chemistry: Assessing KR&R Technologies. In *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning*, pages 488–497, 2004.

[2] D. Brickley and A. Miles. SKOS Core Vocabulary Specification. W3C working draft, W3C, November 2005.

[3] J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson. Jena: Implementing the semantic web recommendations, 2003.

[4] CiteULike. http://www.citeulike.org/, 2007. (accessed 2007-02-11).

[5] Connotea. http://www.connotea.org/, 2007. (accessed 2007-02-11).

[6] del.icio.us. http://del.icio.us/, 2007. (accessed 2007-02-11).

[7] FURL. http://www.furl.net, 2007. (accessed 2007-02-11).

[8] Google Web Toolkit. Google Web Toolkit - Build AJAX apps in the Java language. http://code.google.com/webtoolkit/, 2007. (accessed 2007-01-25).

[9] M. Hepp. Possible Ontologies: How Reality Constraints Building Relevant Ontologies. *IEEE Internet Computing*, 2006.

[10] Hibernate. Relational Persistence for Java and .NET. http://www.hibernate.org/, 2007. (accessed 2007-02-11).

[11] A. Hotho, R. Jschke, C. Schmitz, and G. Stumme. BibSonomy: A Social Bookmark and Publication Sharing System. In A. de Moor, S. Polovina, and H. Delugach, editors, *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, Aalborg, Denmark, July 2006. Aalborg University Press.

[12] HSQLDB. Lightweight 100% Java SQL Database Engine. http://hsqldb.org/, 2007. (accessed 2007-02-11).

[13] M. Krötzsch, D. Vrandecic, and M. Völkel. Semantic MediaWiki. In I. F. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, *International Semantic Web Conference*, volume 4273 of *Lecture Notes in Computer Science*, pages 935–942. Springer, 2006.

[14] Lucene. Apache Lucene is a high-performance, full-featured text search engine library written entirely in Java. http://lucene.apache.org/, 2007. (accessed 2007-02-11).

[15] S. Schaffert. IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. In *WETICE*, pages 388–396. IEEE Computer Society, 2006.

[16] Tomcat. Official reference implementation of for the Java Servlet and JavaServer Pages technologies. http://tomcat.apache.org/, 2007. (accessed 2007-02-11).