The CKC challenge: exploring tools for collaborative knowledge construction

IEEE Intelligent Systems
www.computer.org/intelligent

The CKC Challenge: Exploring Tools for Collaborative Knowledge Construction

Natalya F. Noy, Abhita Chugh, and Harith Alani

Vol. 23, No. 1
January/February 2008

Editor: Steffen Staab
University of Koblenz-Landau
staab@uni-koblenz.de

The CKC Challenge: Exploring Tools for Collaborative Knowledge Construction

Natalya F. Noy, *Stanford University*
Abhita Chugh, *Yahoo!*
Harith Alani, *University of Southampton*

Web 2.0's great success is fueled mainly by an infrastructure that lets users easily create, share, tag, and connect content and knowledge. In general, the knowledge created in today's applications for collaborative tagging, folksonomies, wikis, and so on is mostly unstructured: tags or wiki pages don't have semantic links between them and usually aren't related in any structured form. By contrast, ontologies, database schemas, and taxonomies usually contain explicit definitions of, and links between, their components, often with well-defined semantics.

A new generation of tools supports the integration of Web 2.0 and Semantic Web approaches. Some of these tools—such as Semantic MediaWiki (http://meta.wikimedia.org/wiki/SemanticMediaWiki), BOWiki, (http://onto.eva.mpg.de/bowiki), and Platypus Wiki (http://platypuswiki.sourceforge.net)—provide wiki extensions for creating semantic links between pages. Other tools let users organize tags in some semantic structure, and fully fledged ontology editors such as pOWL support the distributed and collaborative development of ontologies. Commercial tools such as Freebase (http://freebase.com) are also entering the field.

Most of these tools are in early development—the collaborative-knowledge-construction field is in its infancy. Few, if any, user studies outline what users expect from such tools and what does or doesn't work. So, we organized the Collaborative Knowledge Construction Challenge. The CKC Challenge let users try different tools and provide feedback to help us assess the state of the art.

The CKC Challenge

We ran the challenge under the auspices of the Workshop on Social and Collaborative Construction of Structured Knowledge, held at the 16th International World Wide Web Conference. We solicited tool demonstrations for the workshop. Those tools that were domain independent, dealt with structured knowledge, and let users work collaboratively became part of the CKC Challenge. (We use "structured knowledge" to describe anything from a simple hierarchy of tags to a fully axiomatized OWL ontology.)

Our goal wasn't to compare the tools but to understand requirements and alternatives for developing such tools. The developers of the tools we assessed actively supported the challenge and were keen to observe user experiences and obtain detailed feedback.

The tools

Table 1 describes the six tools in the challenge: BibSonomy,[1] Collaborative Protégé, DBin,[2] Hozo,[3] OntoWiki,[4] and SOBOLEO.

The tools' functionality and focus differed, thus letting challenge participants try out a variety of features (see table 2). All six tools let you create hierarchical information. However, Collaborative Protégé, DBin, and OntoWiki focus on editing ontologies, while Hozo focuses on building and integrating ontology modules and detecting resulting conflicts. BibSonomy and SOBOLEO are primarily for tagging Web resources and creating taxonomies by specifying relationships between tags. Collaborative Protégé and SOBOLEO let users chat and create discussion threads, and Collaborative Protégé and OntoWiki let them rate ontology components and add comments to ontology components. BibSonomy separates a user's personal space from the shared space of all users; in the other tools, users share all information. Collaborative Protégé and OntoWiki maintain a log of changes made to the ontology by different users and make it available during editing. Hozo supports asynchronous knowledge construction and lets users lock ontology components before editing them.

Clearly, the tools target different scenarios for collaborative knowledge construction. In Collaborative Protégé and Hozo, for example, the final product is the ontologies themselves. In BibSonomy and SOBOLEO, users annotate Web resources with tags. The tag hierarchies in BibSonomy and SOBOLEO are byproducts of collaborative anno-

1541-1672/08/$25.00 © 2008 IEEE
Published by the IEEE Computer Society

Table 1. The tools that participated in the Collaborative Knowledge Construction Challenge.

Tool	Description	Developing institution
BibSonomy[1] (www.bibsonomy.org)	A Web-based social resource-sharing system that lets users organize, tag, and share bookmarks and bibliographic entries. Users can create relations between tags. For example, **programming** ← **Java** indicates that **programming** is a more general tag than **Java**. So the system should include resources tagged with **Java** when the user searches for **programming**.	University of Kassel
Collaborative Protégé (http://protege.stanford.edu)	An ontology and instance editor that supports concept history, discussions, and comments to both ontology components and ontology changes. For example, users can suggest or explain changes, create proposals for changes, and vote on them.	Stanford University
DBin[2] (http://dbin.org)	A peer-to-peer application that lets users collaboratively edit knowledge bases. The CKC *brainlet* (a domain-specific user interface generated by DBin from a collection of components) includes ontology-editing and instance-acquisition components. The system provides provenance information for classes and properties.	Università Politecnica delle Marche
Hozo[3] (www.hozo.jp)	An ontology editor that lets users asynchronously develop ontologies that are subdivided into multiple interconnected modules. A user can check out and lock a specific module, edit it locally, and then check it back in. Hozo uses dependencies declared between modules to identify conflicts. Users can view the list of changes and accept or reject them.	Osaka University
OntoWiki[4] (http://3ba.se)	A Web-based ontology and instance editor that provides such capabilities as history and ratings. OntoWiki provides different views on instance data (for example, a map view for geographical data or a calendar view for dates).	University of Leipzig
SOBOLEO	A Web-based system that lets users assign tags to Web resources and organize the tags in a hierarchy. Users can search through the annotated Web resources using concepts from the taxonomy.	FZI Research Center for Information Technology

tation and are used mainly to search and browse the annotated resources. DBin and OntoWiki focus on knowledge acquisition, letting users create and share knowledge bases and enter new instance data using customized interfaces.

Preparations

Before the challenge, we used each tool to enter some bootstrapping information about the CKC workshop (papers and first authors). By doing so, we effectively tested each tool and its suitability to the task. Furthermore, for the tools that provided discussion, annotation, or rating facilities, we demonstrated these features.

This bootstrapping period was essential because it let us test the tools in the challenge setting to ensure we set them up correctly. We identified tool features that were difficult to use or find and asked the tool developers to add clarifications to a how-to page for the challenge.

Participation

The challenge ran for two weeks, and anyone could participate. We invited users to enter information on their research domains, because we thought that having them enter information about themselves and their institutions and papers—information close to their heart—would encourage participation. More specifically, we asked users to construct structured knowledge for a hypothetical portal for research information. We suggested that users start by capturing the information about their own research topics, research groups, publications and conferences, or other events in which they were involved. Users were free to expand the knowledge base in any direction. Because the participating tools differed in their expressive power and the type of information they supported, the exact information that users could enter differed for each tool.

To further encourage participation, we announced a competition for the most active user.[5] We collected each tool's activity information daily and aggregated it on the challenge Web site so that users could track their activity levels (similar to a computer game). Different tools computed the "number of facts entered" differently, so adding information about a paper in one tool might count as a different number of operations in another tool. However, because the competition was rather informal (and the prizes symbolic), this discrepancy wasn't a major concern.

Also, our scoring system explicitly encouraged users to try more than one tool: making 30 changes in a single tool counted less than using three tools and making 10 changes in each. Edits in the second and subsequent tools (in terms of the user's activity level) were more valuable than edits made using the preferred (most used) tool.

Several users were clearly racing each other, entering a large amount of information—particularly toward the challenge's end. The two most active users entered more than 2,000 assertions each over the two weeks.

We asked users to complete a feedback form for each tool they tried as well as a general form. On the tool-specific form, the feedback included which features they liked, what needed improvement, which content type was easy to enter, and which content type wasn't supported. For the more general feedback, we asked users to describe their ideal tool for collaboratively constructing structured knowledge.

Results

Over the two weeks, the CKC Challenge Web site had visitors from over 60 countries. User activity peaked during the last three days. Forty-nine users registered for

Table 2. Challenge tools' features.

	BibSonomy	Collaborative Protégé	DBin	Hozo	OntoWiki	SOBOLEO
Hierarchy of concepts	X	X	X	X	X	X
Properties		X	X	X	X	
Instances of concepts in the hierarchy (includes tags)		X	X	X*	X	X
Comments on ontology components		X	X		X*	
Ratings		X			X	
Asynchronous editing of ontology modules				X		
Personal space	X			X		
History of changes		X	X*	X (module level)	X	
Discussion		X	X*			X
Instant chat		X				X
Content creation using an integrated browser button	X					X
Web browser interface	X				X	X

* Features available in some version of the tool but not in the version used for the challenge.

the challenge, of which 33 actively participated. Over 50 percent of those 33 tried only one tool. We received 36 feedback forms: 31 tool-specific and five general forms.

SOBOLEO had the highest user activity, with 65 percent of the active users trying it (see figure 1). Users seemed to prefer one type of system over another. The tools represented a range of functionality—SOBOLEO and BibSonomy focus on annotating resources, such as Web pages and bibliography entries, while the other tools focus on building structured data and ontologies. So it makes sense that users who tried SOBOLEO also tried BibSonomy, and users who tried Collaborative Protégé also tried OntoWiki. For those who used more than two tools, the two tools they used most were of the same type.

Users largely appeared to follow the representation patterns that the challenge organizers established when entering the initial data. This phenomenon, known as cumulative advantage,[6] is probably a necessary evil in this setting. If users had no initial data on which to model their own input, they might not have entered anything. Yet, entering the initial data seems to have constrained what the users entered.

As figure 1 shows (and user feedback confirmed), download was a big impediment. The two least-used tools, DBin and Hozo, required download and installation. Furthermore, participants used the tools only for the sake of this challenge, rather than for any of their usual work. Perhaps if there was a pressing need for a particular tool, more people would be willing to download it. However, several users pointed out explicitly that a Web interface (and not, say, an applet, such as Collaborative Protégé) was ideal for such tools.

We also noted that there was little discussion, even in the tools that enabled it (Collaborative Protégé and SOBOLEO). Perhaps this lack of discussion was because the domain was too straightforward—there simply wasn't much to discuss or argue about. Another possible explanation is that users had no real use for the resulting knowledge and thus weren't bothered by some imperfections.

What we learned

We base this discussion on information from feedback forms we received and from the discussion of the challenge at the CKC workshop.

The first thing we discovered was that "collaboration" means different things to different people. To many, a tool supports collaborative development if it simply provides distributed users access to the same ontology or knowledge base. To others, a tool supports collaborative development only if it supports annotations from different users, provenance, ways to reach consensus, and so on. Clearly, we as a community still don't have a good idea of what these tools should do or how they should work. Furthermore, everyone agreed that the perfect tool doesn't yet exist.

Here we discuss some of the things we learned that might help improve these tools and further the field.

Collaborative workflows and scenarios

Different scenarios require different workflows and hence different tool support. A group of users developing an ontology in the context of a specific project will have different requirements than an open community developing a lightweight taxonomy that anyone can edit.

The participants expressed different expectations from the tools for adding structure to folksonomies than from tools for collaboratively developing full-fledged ontologies. Everyone seemed to agree that there should be several of such tools to satisfy the varying requirements and settings of different collaborative-development scenarios. However, it will be crucial for the various tools to interoperate by enabling content to be imported to and exported from one another. Users praised tools with these functionalities.

The spectrum of collaboration scenarios that the tools must eventually support is quite broad. In some cases, tools should support specific protocols for making changes—so some users can propose changes, others can discuss and vote on them, but only authorized users can actually make the changes. At the other end of the spectrum are settings where anyone can make immediate changes.

So, tools need to support different mechanisms for building consensus, depending on whether the environment is open or controlled. Although none of the challenge tools specifically supported any consensus-building protocol, several provided some technical means to support such a process. These mechanisms included rating, voting, and discussion.

Expressive power

Different tasks and settings require different levels of expressive power in tools. For instance, although a simple taxonomy might be sufficient for organizing tags for shared annotation of Web resources, users developing medical terminology might need a more expressive ontology. So, not all communities and settings will require the sophisticated features some tools provided. Even though some tools supported ontology development, none provided capabilities for using an expressive ontology language such as OWL. This lack of expressive power didn't bother anyone. In fact, some suggested that it might be difficult to collaboratively edit expressive ontologies.

Trust, credibility, and provenance

Naturally, open environments in which anyone can join the editing process will need more advanced support for establishing user trust and credibility. Given the task's complex nature, poor entries could result from not only malicious intent but also a simple lack of experience or knowledge.

One possible approach is to measure users' reputations on the basis of how the community votes or rates their edits and change proposals. Similarly, support for and easy access to provenance information is critical in distributed collaborative projects. Users must be able to see who made the changes and when, read a comment by the change's author, understand the state of the knowledge base when the change was made, access concept and change histories, and so on.

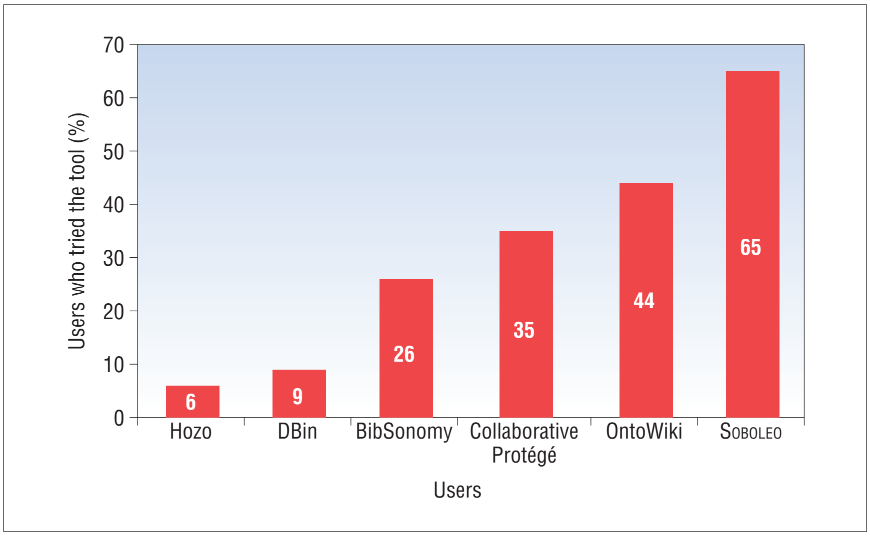

Figure 1. Of the 33 active participants, the most (65 percent) tried SOBOLEO, followed by OntoWiki (44 percent) and Collaborative Protégé (35 percent).

Ontology comprehension

Ontology visualization and explanation—something at which ontology development tools never seem to fully succeed—becomes an even bigger issue in the collaborative setting.[7] Users need to understand and modify ontology elements developed by others, so having visual aids and other features to increase ontology comprehension is critical. In fact, understanding the ontologies and the more sophisticated tools' features might be one of the field's bigger challenges. Consequently, it's not surprising that users liked the visual aids that some of the tools provided, such as Hozo's interactive graphs and OntoWiki's map and calendar widgets.

User interface design

Several participants noted that a good user interface is critical. Users mentioned that having numerous features is useless if they're hard to find or use. In general, participants agreed that having a Web interface to the tool becomes increasingly important, and new technologies, such as AJAX (Asynchronous JavaScript Technology and XML), let tools support a wide range of sophisticated user interface features in this setting. Some users suggested building browser plug-ins to speed up inserting and accessing knowledge from some of the lightweight tools (such as BibSonomy and SOBOLEO).

Personal and shared spaces

A controversial topic that elicited emotional response from both sides was whether tools should support personal spaces for a user's own structures and data (like BibSonomy) or immediately share everything with the community (Collaborative Protégé, OntoWiki, and SOBOLEO). In a third (intermediate) case, a user can develop data asynchronously in his or her personal space before posting it to the world (Hozo).

Ultimately, there was no agreement on which model works best. Different settings will probably necessitate different boundaries between personal and shared space. Personal views, which none of the tools currently implement, might address some aspects of this issue. Users might often want to restrict the view to reflect their personal decisions or their trust network, for example.

Mechanisms for building consensus

Collaborative Protégé and OntoWiki let users rate concepts and individuals. However, users didn't find such ratings useful unless they included comments and explanations. Other users suggested using conventional chat software (such as Yahoo! or MSN) for the discussions instead of implementing tool-specific chat services. There were some general concerns that discussions and voting techniques in general could slow down editing if tools demand a consensus on every editing action before it can take effect.

It remains to be seen whether complex ontologies are even amenable to collaborative editing. Some users felt that the space is too complex for this type of work. As one participant pointed out, "It's hard enough to agree on ontological decisions when people are sitting at the same table. I doubt that any nontrivial ontology can be built using only asynchronous communication such as the discussion tool."

At the same time, the tool-supported environment might help by providing a record of changes and comments, recording the discussion and directly linking it to the ontology components, and helping frame the discussion's context.

One main result of this challenge was the realization of how the different tools can learn from each other and exchange features. For example, several users suggested adding BibSonomy and SOBOLEO's Web-resource-annotation feature to core ontology-editing tools such as Collaborative Protégé and Hozo. The developers of BibSonomy and SOBOLEO are now integrating these two tools to bridge the gap between BibSonomy's tag-based approach with SOBOLEO's more ontology-based approach.

More information on the challenge tools and results appear in the workshop proceedings[8] and online at http://km.aifb.uni-karlsruhe.de/ws/ckc2007/challenge.html. ■

Acknowledgments

Domenico Gendarmi (University of Bari, Italy) was the most active challenge participant. Alireza Tajary (Amirkabir University, Iran) was runner-up. Meredith Taylor (Macquarie University, Australia) gave the most insightful feedback. We thank the tool developers for their support and for feedback on an earlier draft of this article. We're most grateful to the challenge participants. Finally, we thank the workshop's co-organizers: Gerd Stumme, Peter Mika, York Sure, and Denny Vrandeĉić.

References

1. A. Hotho et al., "BibSonomy: A Social Bookmark and Publication Sharing System," *Proc. Conceptual Structures Tool Interoperability Workshop at the 14th Int'l Conf. Conceptual Structures*, Aalborg Univ. Press, 2006, pp. 87–102.
2. G. Tummarello, C. Morbidoni, and M. Nucci, "Enabling Semantic Web Communities with DBin: An Overview," *Proc. 5th Int'l Semantic Web Conf.* (ISWC 06), LNCS 4273, Springer, 2006, pp. 943–950.
3. E. Sunagawa et al., "An Environment for Distributed Ontology Development Based on Dependency Management," *Proc. 2nd Int'l Semantic Web Conf.* (ISWC 03), LNCS 2870, Springer, 2003, pp. 453–468.
4. S. Auer, S. Dietzold, and T. Riechert, "OntoWiki—A Tool for Social, Semantic Collaboration," *Proc. 5th Int'l Semantic Web Conf.* (ISWC 06), LNCS 4273, Springer, 2006, pp. 736–749.
5. B. Good et al., "Fast, Cheap and Out of Control: A Zero Curation Model for Ontology Development," *Pacific Symp. Biocomputing*, vol. 11, 2006, pp. 128–139.
6. D. Watts, *Six Degrees: The New Science of Networks*, Vintage, 2004.
7. A. Gibson, K. Wolstencroft, and R. Stevens, "Promotion of Ontological Comprehension: Exposing Terms and Metadata with Web 2.0," *Proc. Workshop Social and Collaborative Construction of Structured Knowledge at WWW 2007*, CEUR, 2007, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-273.
8. *Proc. Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th Int'l World Wide Web Conf.*, N.F. Noy et al., eds., CEUR, 2007, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-273.

Natalya F. Noy is a senior research scientists at Stanford University. Contact her at noy@stanford.edu.

Abhita Chugh is a software engineer at Yahoo! Contact her at abhita@gmail.com.

Harith Alani is a senior research fellow at the University of Southampton. Contact him at ha@ecs.soton.ac.uk.