# Survey of tools for collaborative knowledge construction and sharing

Gianluca Correndo, Harith Alani
School of Electronics and Computer Science
University of Southampton
SO17 1BJ, Southampton, United Kingdom
gc3@ecs.soton.ac.uk, ha@ecs.soton.ac.uk

## Abstract

*The fast growth and spread of Web 2.0 environments have demonstrated the great willingness of general Web users to contribute and share various type of content and information. Many very successful web sites currently exist which thrive on the wisdom of the crowd, where web users in general are the sole data providers and curators. The Semantic Web calls for knowledge to be semantically represented using ontologies to allow for better access and sharing of data. However, constructing ontologies collaboratively is not well supported by most existing ontology and knowledge-base editing tools. This has resulted in the recent emergence of a new range of collaborative ontology construction tools with the aim of integrating some Web 2.0 features into the process of structured knowledge construction. This paper provides a survey of the start of the art of these tools, and highlights their significant features and capabilities.*

## 1. Introduction

The transformation of the web from a mere collection of documents to a queryable knowledge-base (KB) is one of the most prominent targets of Semantic Web (SW, [2]).

To help reach this goal, knowledge repositories need to publish semantic representations of their data models to enable other machines to understand and query their content. To this end, much research and development has focused on building tools and capabilities for ontology and KB construction. However, most of such tools have been designed to be used by individuals or small teams that are geographically co-located. Support for distributed teams to remotely and continuously collaborate on building ontologies and knowledge repositories has been somewhat lacking in the SW domain.

Defining an ontology for representing data semantics is usually a costly and time consuming task. Furthermore, knowledge evolves over time and this is reflected in costly reworks. If an ontology is meant to represent the views of a specific community and support their knowledge needs, then it might be sensible to give this community the power to maintain this ontology [8].

Some ontologies need to be agreed upon by the user community, and reaching this agreement must be supported by tools and methodologies to allow users to express their views and opinions freely, using for example tags to annotate certain objects or highlight some issues.

The rise of social Web 2.0 applications has demonstrated how general web users can actively contribute and share all sorts of data and information, such as images, videos, bookmarks, opinions, diaries and experiences, etc. Mapping this functionality to the SW means giving more support to users to dynamically and collaboratively build ontologies add semantics to data, discuss and share views and suggestions, etc. Good and colleagues (see [5]) showed how SW users can successfully collaborate to negotiate and build good quality ontologies when provided with a tool that supports such activities. In the following we will describe the main features and capabilities of a number of tools that offer various levels of collaboration with the goal of creating structured knowledge.

## 2. Collaborative Knowledge Construction

Since the early nineties collaboration has been seen as a key argument for constructing consensual knowledge around a domain and a great effort has been spent in discovering enabling technologies for facilitating such consensus reaching. Several tools have emerged in the past few years which support various kinds and levels of semantic knowledge creation. This knowledge may vary from simple bookmarking information, to instance data, to complex OWL ontologies. In this paper we first have a look at some of the early developments in the field and then survey the tools presented at the workshop on Social and Collaborative Construction of Structured Knowledge (http://www2007.org/workshop-W7.php) at the $16^{th}$ International World Wide Web conference.

## 2.1. Related Work

The distributed editing and maintenance of formal artifacts, like knowledge bases or ontologies, have posed some challenges to computer science with a number of issues: how to better orchestrate collaborative efforts, how to assure global consistency, how to reach consensus, and how to enhance understanding. The first approaches applied a *divide et impera* philosophy by dividing domains of interest into modules with well established dependencies.

Ontolingua [4] was one of the first tools for authoring formal ontologies with a collaborative flavour. The tool manages shared sessions where users can collaboratively edit a set of ontologies. Ontoligua supports user groups and different access rights, similar to a traditional distributed filesystem. All users participating in an ontology building session in Ontolingua are notified of any changes inserted by other users. Users can reuse and share ontology modules, thus creating a collaborative environment. The former language used in Ontolingua is KIF, a very expressive logic, and the tool provided a consistency checking service.

Ontolingua provided valid solutions for achieving two of the four goals stated above, **orchestration** and **global consistency**, but, since different subdomains were treated in different modules, it did not address in depth how to express different conceptualisations and modelling issues. Some later approaches tried to augment models with argumentation about the validity of proposals or discussing on their faults. One tool that integrated argumentations in proposals is $Co_4$ [3] which aimed to increase consensus and cooperation in KB creation by having a protocol for managing discussions.

In $Co_4$, each user maintains a local KB that remains unshared until the user decides that it is mature enough to be shared, totally or in part, with the community. The sharing activity is undertaken by means of change requests submitted by users to the approval of the group who maintain a consensual KB. The submissions are managed with axioms that the group can "accept", "reject" or subject to "revision", in regard of providing suggestions for improving the proposal.

## 2.2. BibSonomy

BibSonomy [6] is a social tagging system developed by the University of Kassel and the L3S research institute of Hannover. BibSonomy helps in organising and sharing resources such as bookmarks and publications, integrating capabilities separately provided by previous social tagging systems. The collection of tags is a weak notion of controlled vocabulary where 'control' is performed by the community itself. BibSonomy introduced the possibility to add an informal notion of *isa* relation between tags in order to

provide a notion of taxonomic structure that can be helpful when retrieving information. The idea is that tags that are more general than the ones used for a resource are implicitly added to the set of tags and used when searching for information if required. Using these relations, the user is not forced to apply all the possible tags to resources but only the most specific ones.

The interaction provided by BibSonomy is similar to common social bookmarking systems. A user submitting a resource (eg reference, web page) to BibSonomy will be recommended a number of tags that were given to the same resource by others, which the user may choose to reuse, or add new ones. Bookmarked resources can be made private or shared among the whole community. When shared, the bookmark and all associated tags can be searched by others and used by the system for future tag recommendations. Users can formulate groups where they can limit the sharing of tags and resources to just the other members of the group.

## 2.3. DBin

DBin [10] is an application that allows a group of users to share structured content using RDF as a content language and a P2P network as a content provisioning infrastructure. The tool supports the most well known query languages for RDF (i.e. RDQL, RQL and SeRQL) and decorates each information provided to the group with trust related information. The main services that DBin provides can be summarised as follow:

- Application container: the tool provides a set of reusable widgets that handle RDF data (e.g. for visualising maps or URLs).

- Provenance certificates: every content provided by a user to the community using DBin is certified by a signature that uniquely identifies the user.

- RDFGrowth: a P2P algorithm that allows users and communities to specifically define the items that are of interest in order to receive only the relevant updates.

## 2.4. Hozo

Hozo [7] is a server based tool developed by the university of Osaka for supporting the development of ontologies in a distributed environment. The main focus of the tool is to provide consistency among inter-dependent ontologies in a scenario where they are asynchronously developed by different users. In order to provide such global consistency, the tool provides basically two capabilities: dependency management and consistency harmonization.

Like many version control systems, Hozo allows to lock and unlock modules, that in this case are ontologies present

in the system, in order to prevent any modifications from other users. The state of the source ontologies is broadcasted to all users along with their update status in order to let users know, which ontologies are in rework and what modifications have affected the focused concepts. Modifications applied to the ontology entities (concepts and slots) are of three kinds: added, modified and deleted, and the user can accept or reject each modification.

## 2.5. OntoWiki

OntoWiki [1] is a web based tool for collaborative editing of information maps that supports a number of visualisation widgets for different kind of instance data. The aim of OntoWiki is to provide an intuitive and collaborative tool for the editing of RDF content that supports knowledge engineering in a distributed (web based) environment. The user can browse the taxonomy tree of the concepts. Once a concept is selected, the user can see its instances. Once an instance is selected, the user can then choose a visualisation for it or edit the RDF structure to add additional properties or values to the instance. OntoWiki provides a text search facility on literal values, supported by other search filtering options on frequency, instance groups, facets, etc.

OntoWiki also provides a number of features for enhancing social interaction when developing KBs. The features implemented follow the Wiki paradigm where the system is seen as a common space open to contributions where users can collaborate to build information and to correct mistakes. All these features are intended to lower the cost of creating and exchanging structured metadata and to set up an effective scenario of collaboration. OntoWiki's main features can be summarised as follows:

- Change tracking: KB changes are tracked and the users can subscribe to RSS feeds for being informed

- Commenting: all the changes can be commented on, annotated and rated by users in order to discuss and agree on their usefulness

- Popularity: all accesses to the KB are logged and instances can be rated (e.g. for originality, quality and presentation) by users in order to measure instance popularity

- Provenance: the system tracks contributions from users to highlight whose contributions are added to the common task

## 2.6. Collaborative Protégé

Collaborative Protégé [9] is a prototypical extension of the Protégé ontology editor from Stanford Medical Informatics department for addressing social editing of ontologies. The tool adopts a frame-like representation paradigm for editing ontologies, but supports axioms and many formalisms for formal ontology description (RDF(S), OWL, XMI, N3). The basic version of the tool already supports a server based editing of the ontologies, based on Java RMI (Remote Method Invocation), where the concept of predefined users and groups is used to define access policies to the ontologies. In this prototypical extension, the users can make changes on a shared model and see in real time the changes made by other users.

For empowering a community to build an agreed complex artifact, such as an ontology, Protégé has provided some features for addressing the discussion between users and conflict resolution by means of annotations that describes changes, proposals, votes, advice, comments etc. Using such additional annotation concepts, the user can provide structured information focused on the desired evolution of the ontology and discuss it with the rest of the community. Moreover, the tool allows a basic rating capability for annotating in order to clearly identify the more active topics and instant messaging for connecting users with same interests on the model.

## 2.7. SOBOLEO

SOBOLEO [11] is a web based collaborative tool for the engineering of taxonomies developed by the Research Center for Information Technologies of the University of Karlsruhe. The tool allows users to mantain a shared taxonomy, using SKOS concepts (http://www.w3.org/2004/02/skos/), and to use such concepts for tagging internet resources. This integration of functionalities has been devised for bridging the gap between the creation of shared vocabularies and their use in organising internet resources.

The tool has a taxonomy editor for changing the hierarchical definition of the concepts (no slots or axioms) and such changes are not supervised by a central authority. The editor allows users to send instant messages to users that are currently editing the taxonomy. Using concepts from the mantained taxonomy, the user can tag internet resources, and the tool will prompt him/her with concepts used by others on same resources. SOBOLEO provides resource retrieval features as well. Using concept from the managed ontology as keywords, the user can search tagged resources (refining the search using broader and narrower concepts) or browse the taxonomy to look for tagged documents.

## 3. Comparison

Table 1 compares the main features provided by the tools surveyed above with respect to model editing and the degree of formality of the model itself. It is important to know if the semantics of the model have a strong formal basis or not as this will influence the use of the tool for a particular task.

Moreover, the table denotes, for each tool, the possibility to define relations between concepts other than the classic **isa** and indicates whether it is possible to create instances from the used ontologies. Finally, the last column shows whether the tools allows using concepts for tagging external resources like bookmarks, pictures etc.

**Table 1. Comparison of model editing features**

| Tool | Formality | Properties | Instances | Tags |
|------|-----------|------------|-----------|------|
| **BibSonomy** | Uncontrolled Vocabulary | | | ✓ |
| **DBin** | Formal Ontology | ✓ | ✓ | ✓ |
| **Hozo** | Formal Ontology | ✓ | ✓ | |
| **OntoWiki** | Formal Ontology | ✓ | ✓ | |
| **Collab. Protégé** | Formal Ontology | ✓ | ✓ | |
| **SOBOLEO** | Taxonomy | | | ✓ |

Table 2 compares the features provided by the tools for driving discussions and negotiations about the model among the community. The first feature taken into account is the possibility to supervise the changes on the model having an effective control over its evolution. An entry in the second column indicates that the tool provides the ability to add argumentation to the evolution of the model in order to reach consensus or to discuss issues. The last feature taken into account for comparing the tools is whether the users can be connected by means of social tools with other users of the community in support of the task (e.g. chat, instant messaging, RSS).

**Table 2. Comparison of model evolution features**

| Tool | Supervision | Argumentation | Social tools |
|------|-------------|---------------|--------------|
| **BibSonomy** | | | ✓ |
| **DBin** | | | |
| **Hozo** | | | |
| **OntoWiki** | | | ✓ |
| **Collab. Protégé** | ✓ | ✓ | ✓ |
| **SOBOLEO** | | | ✓ |

## 4. Conclusions

This paper presented a brief survey of tools submitted to the workshop for Social and Collaborative Construction of Structured Knowledge which aimed at testing the state of the art of such tools for collaborative knowledge construction to find out what users expect from such tools, and how they should evolve in the future. The tools surveyed in this paper support various tasks, such as tag-based taxonomy building, KB creation, and ontology modelling. They also vary greatly in terms of the functionalities they provide for tagging, voting, discussion support, change control, visualisation, etc. This paper highlighted and compared the main features of these tools to help reach a better understanding of what is currently available.

## 5. Acknowledgment

## References

[1] S. Auer, S. Dietzold, J. Lehmann, and T. Riechert. Ontowiki: A tool for social, semantic collaboration. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007*, Banff, Canada, 2007.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.

[3] J. Euzenat. Building consensual knowledge bases: Context and architecture. In N. Mars, editor, *Towards Very Large Knowledge Bases - Proceedings of the KB&KS '95 Conference*, pages 143–155, 1995.

[4] A. Farquhar, R. Fikes, and J. Rice. The ontolingua server: A tool for collaborative ontology construction, 1996.

[5] B. M. Good, E. M. Tranfield, P. C. Tan, M. Shehata, G. K. Singhera, J. Gosselink, E. B. Okon, and M. D. Wilkinson. Fast, cheap and out of control: A zero curation model for ontology development. In *Pacific Symposium on Biocomputing*, pages 128–139, 2006.

[6] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. BibSonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, 2006.

[7] K. Kozaki, E. Sunagawa, Y. Kitamura, and R. Mizoguchi. Distributed and collaborative construction of ontologies using hozo. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007*, Banff, Canada, 2007.

[8] N. Shadbolt, T. Berners-Lee, and W. Hall. The semantic web revisited. *Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications]*, 21(3):96–101, 2006.

[9] T. Tudorache and N. Noy. Collaborative Protégé. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007*, Banff, Canada, 2007.

[10] G. Tummarello and C. Morbidoni. Collaboratively building structured knowledge with dbin: from del.icio.us tags to an 'rdfs folksonomy'. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007*, Banff, Canada, 2007.

[11] V. Zacharias and S. Braun. Soboleo – social bookmarking and lightweight engineering of ontologies. In *Proc. WWW 2007 Workshop on Social and Collaborative Construction of Structured Knowledge*, Banff, Canada, May 2007.