Fostering knowledge evolution through community-based participation

Domenico Gendarmi
University of Bari
Dipartimento di Informatica
Via E. Orabona, 4 - 70125 Bari
+390805442286

gendarmi@di.uniba.it

Fabio Abbattista
University of Bari
Dipartimento di Informatica
Via E. Orabona, 4 - 70125 Bari
+390805443298

fabio@di.uniba.it

Filippo Lanubile
University of Bari
Dipartimento di Informatica
Via E. Orabona, 4 - 70125 Bari
+390805443261

lanubile@di.uniba.it

ABSTRACT

The ontology development process is typically led by single or small groups of experts, with users mostly playing a passive role. Such an elitist approach in building ontologies hinders the primary purpose of large-scale knowledge sharing. Collaborative tagging systems have emerged as a new web annotation method proving appealing features in fostering users to collaboratively organize information through their own metadata. Collaborative tagging shifts the creation of metadata for indexing web resources, from an individual professional activity to a collective endeavor, where every user is a potential contributor. In this paper we introduce an approach to knowledge evolution which aims to exploit the ability of collaborative tagging in fostering community members participation to move forward an initial knowledge structure. We present user scenarios about how subscribers of a scientific digital library might play the role of knowledge organizers through personal organization and sharing of citations of interest.

Categories and Subject Descriptors

H.3.5 Online Information Services, H.3.7 Digital Libraries, H.5.3 Group and Organization Interfaces.

General Terms

Design, Human Factors.

Keywords

Community, knowledge evolution, collaborative tagging.

1. INTRODUCTION

Knowledge is strongly tied up with cognitive and social aspects, as the management of knowledge occurs within a tangled structured social context. Human and social factors involved in the development and exchange of knowledge have a heavy impact on the design of knowledge management supporting systems [16]. Such a collaborative knowledge construction process takes place when multiple participants contribute to the growth of interpretations on a shared information base, simultaneously extended by information seeking and transformations [15].

In order to help community members constructing knowledge in their own personal perspectives while also negotiating shared understanding, two needs have to be addressed: First, people need

ways to find and work with information that matches their personal needs, interests, and capabilities. Then people need to bring together their individual knowledge to build a shared understanding and collaborative outcomes [14]. This can be accomplished by the Semantic Web whose main goal is to enable computers and people to work in cooperation [1].

Ontologies play a relevant role within the Semantic Web vision, because they allow to cope with heterogeneous representations of web resources, providing a common understanding of a domain to be shared among human beings and software agents [6]. The domain model implicit in an ontology can be taken as a unified structure for giving information a common representation and semantic [2]. However the ontology development process is typically led by single or small groups of experts, with users mostly playing a passive role. Such an elitist approach in building ontologies hinders the primary purpose of large-scale knowledge sharing.

The achievement of a widespread participation in the ontology development process is often hampered by entry barriers, like the lack of easy-to-use and intuitive tools for ontology contribution. Barriers to active participation, combined with traditional top-down approaches in building ontologies, force users to conform to an undesirable knowledge representation. Such an imposition weakens common ground and increases the likelihood that the ontology will not be widely used.

Ontologies need to change as fast as the parts of the world they describe [7]. However, changes have to be captured and applied by skilled knowledge engineers, preferably the original creators of the ontology. This is a bottleneck which causes unacceptable delays in the ontology maintenance process.

A reasonable assumption on how to reduce maintenance costs is to spread the burden across users. In fact, given the Web's fractal nature, costs might decrease as ontology users increase in number [13]. Community participation to ontology development has already been identified as a solution to a more complete and up-to-date structured knowledge construction [19]. Other than being group of users with common interests, communities can then be considered as the top layer of the Semantic Web architecture [12].

This paper describes our vision for enabling a community of autonomous users to cooperate in a dynamic and open environment, collectively evolving an initial knowledge structure. Participants can organize some piece of knowledge according to a self-established vocabulary, building up personal taxonomies for searching and browsing through their own information spaces. By sharing portions of their knowledge, users can also create

connections and negotiate meaning with people having similar interests.

The main goals of the proposed approach are: (1) to allow users to organize personal information spaces, starting from a prearranged knowledge structure; and (2) to take advantage of users' contribution for better reflecting the community evolution of a shared knowledge structure.

The rest of the paper is organized as follows. Section 2 provides background information about collaborative tagging systems. In Section 3 we describe our approach to community-based evolution through a specific context, a scientific digital library, and a number of user scenarios. Section 4 summarizes related work that can be seen as complementary to our approach. Finally section 5 draws conclusions and points out some challenges we are going to address in the near future.

2. COLLABORATIVE TAGGING SYSTEMS

One of the major obstacles hindering the widespread adoption of controlled vocabularies is the constant growth of available content which anticipates the ability of any single authority to create and index metadata. In such contexts collaborative tagging represents a potential solution to the vocabulary problem [4].

Collaborative tagging has emerged as a new social-driven annotation method, as it shifts the creation of metadata for describing web resources, from an individual professional activity to a collective endeavor, where every user is a potential contributor.

Figure 1 shows a conceptual model of collaborative tagging, according to UML notation [3], with tags seen as association classes between users and resources. Users can label any resource with whatever tag thought as appropriate and, vice versa, resources can be annotated with any tag by any user. Users are able to share both resources and tags within a community, leading to a network of users, resources and tags with a flat structure and no limits in evolution.
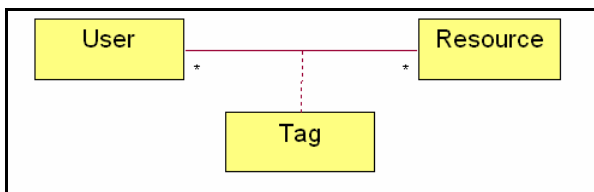

Figure 1. Conceptual model of collaborative tagging

Collaborative tagging systems exhibit other interesting benefits such as their ability in adhering to the personal way of thinking. No forced restrictions on the allowed terms, as well as the lack of syntax to learn can shorten significantly the learning curve. Collaborative tagging systems also create a strong sense of community amongst their users, allowing them to realize how others have categorized the same resource or how the same tag has been used to label different resources. This immediate feedback leads to an attractive form of asynchronous communication through metadata [10]. There is no need to establish a common agreement on the meaning of a tag because it gradually emerges with the use of the system. Marginal opinions

can coexist with popular ones without disrupting the implicit emerging consensus on the meaning of the terms.

The main drawbacks with tags concern semantic and cognitive issues, such as polysemy, synonymy and basic level variation [5]. Polysemy occurs when the same term is used for tags employed with different meanings. The polysemy problem affects query results by returning potentially related but often inappropriate resources. Polysemy is occasionally equalized to homonymy, however polysemous words have different meanings but related senses, while homonyms have multiple, unrelated meanings. Synonymy takes place when different terms are used for tags having the same meaning. Synonymous tags are another source of ambiguity, severely hindering the discovery of all the relevant resources which are available in a tagging system. Polysemy and synonymy represent two critical aspects of a search, as they respectively affect precision and recall, which are typically used for evaluating information retrieval systems.

A further relevant problem, concerning the cognitive aspect of categorization, is the basic level variation of tags. Terms used to describe a resource can vary along a continuum of specificity ranging from very general to particularly specific. Different users can use terms at different levels of abstraction to describe the same resource, leading to a low recall in retrieving resources. Collaborative tagging is also referred to as "folksonomy", originally coined by Thomas Vander Wal who combined the words "folk" and "taxonomy", this term refers to a taxonomy created by common people [17]. However, taxonomies are hierarchical structures of classifications with parent-child relationships among concepts.

While it is well-known that search and retrieval are facilitated by structured subject headings, the tags which form a folksonomy are just flat terms. Besides the previous drawbacks, the lack of a structure is one of the main aspects which weaken severely the information retrieval in a collaborative tagging system.

3. OUR APPROACH TO COMMUNITY KNOWLEDGE EVOLUTION

In this section we lay out our approach for applying collaborative tagging techniques to support the evolution of a knowledge structure adopted for the classification of a wide amount of digital resources.

We first briefly introduce a scientific digital library that we have selected as an application context. Then we present the knowledge evolution process from a user perspective.

3.1 Approach Context

As an illustrative context for our approach, we consider the digital library of the Association for Computing Machinery (ACM).

The ACM Guide to Computing Literature is an index to computing literature from over 3000 publishers, containing over 750,000 citations of books, journal articles, conference proceedings, doctoral and master's theses, and technical reports. Citations can be browsed by publication type, author name, as well as authors' keywords and classification terms from the ACM taxonomy, named The Computing Classification System.

The ACM Guide to Computing Literature is part of the services offered by the ACM Portal. Portal subscribers can create any number of binders, which are personal collections of citations with links to the publication source through the Digital Object

Identifier (DOI) bookmark, and the full text if the citation is published by ACM itself. When creating their binders, users choose whether to keep them private or share them with other selected users or, more generally, the public.

3.2 User Perspective

According to our approach, the interaction process of a user with a digital library can be characterized as a three-step iteration (Figure 2).

1. **Selection**. It involves discovering and choosing a specific citation in the whole repository. This step is already available in a common digital library.

2. **Organization**. It involves creating and structuring a personal information space according to individual interests. This step goes beyond current opportunities because it allows not only to store collections of citations of interest but also to group them using the desired metadata and structure.

3. **Sharing**. It involves making public some selected collections and corresponding metadata in order to support a community knowledge evolution.

To explain how our approach can affect the user experience, afterwards we present a scenario for each step.
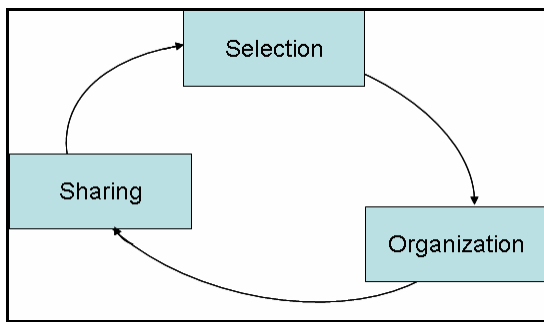

Figure 2. Three-step iteration

3.2.1 Selection

John is an ACM member with a web account on the Portal. As an assignment, he has to write a state of the art about collaborative tagging systems. He is not looking for well-known papers but, rather his goal is to explore the recent bibliography on this specific topic to discover new scientific articles he could find interesting to read.

In order to find citations within the ACM Portal, John has two options: He can perform a search (basic or advanced); otherwise he can browse the repository in several different ways. For example, he can browse through the Guide using index terms of the ACM taxonomy or he can browse through the Digital Library according to the kinds of publications. However, due to the limitations of the current taxonomy in organizing citations, especially for articles about recent topics as collaborative tagging, John prefers to use the search feature.

John performs a simple query, within the Guide, using as keywords the sentence *collaborative tagging*. A list of results showing a set of basic information (e.g. title, authors, publishers, year of publication) for each matching citation is presented to John ordered by relevance. John, then, can select a specific citation to let the system display additional information related to

that article (e.g. abstract, references, index terms, collaborative colleagues). Once explored more in detail some results, John finds as citation of interest the article named "Usage patterns of collaborative tagging systems". John wants to save it into his own personal information space using the "Save this Article to a Binder" feature (Figure 3).
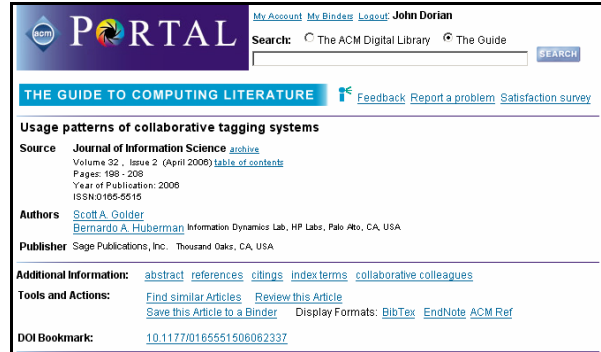

Figure 3. Detailed page of the selected citation

3.2.2 Organization

John now has to choose the name of the binder where saving the selected citation. This name represents the label of a specific category playing the role of a virtual folder where storing a collection of citations. In choosing the name John is supported by a suggestion feature providing a set of potential binder names. In this case some suggested binder names can be *collaborative tagging systems*, *delicious studies* and *social bookmarking analyses*. John chooses to store the citation in a binder named *tagging patterns*.

Saving an article into a virtual personal space is a sign of a real interest for the citation, hence we can assume that John is wishful to provide the metadata he considers most appropriate for annotating the selected citation. However, to avoid burdening John's experience, authoring metadata have to remain as simple as in collaborative tagging systems.

The task assigned to John is just to browse a space of suggested metadata, pointing out the most favorites and eventually proposing new ones. Through the DOI, the system is able to univocally identify the selected citation, and a large set of metadata related to that article can be retrieved from different systems freely available on the web. For example for the selected citation the system could retrieve keywords from ACM, as well as tags from services like CiteULike, Bibsonomy and Connotea (Figure 4).
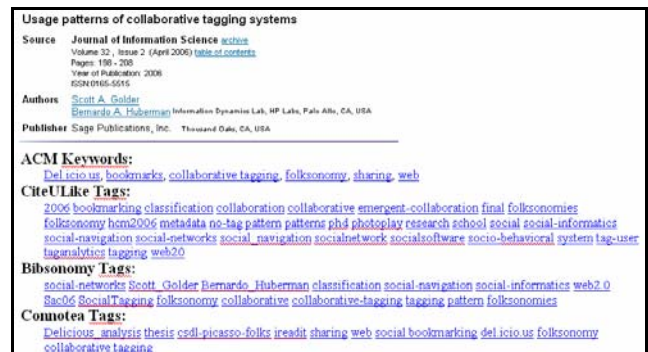

Figure 4. Retrieved metadata of the selected citation

Using a filtering process to discard useless keywords or tags, such as those occurring isolated and group very similar ones, this space of metadata can be normalized in order to help John in the browsing task (Figure 5).

Figure 5. Space of metadata

While browsing, John can select a metadata and, just picking out it, he can state his agreement or disagreement (e.g. Y/N). In this case, browsing the space in Figure 5, John selects *classification* and expresses an agreement with such a term.

Using a lexical resource, such as Wordnet, a searching for possible multiple senses associated to the selected term can be performed. Four senses are retrieved from Wordnet for the noun *classification* and John disambiguates these senses selecting the first one (Figure 6). Furthermore, Wordnet can provide synonyms, hypernyms and hyponyms related to the selected sense (Figure 7). The system can thus map the term chosen by John to a corresponding concept including relationships with other related concepts.
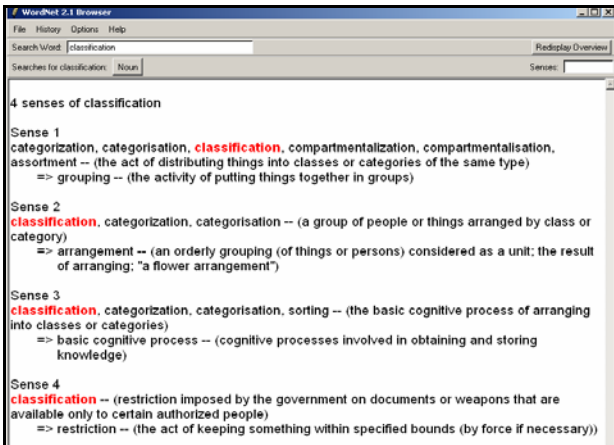

Figure 6. Senses for the term *classification*

John now has to decide the best position, within the ACM taxonomy, where to put the concept corresponding to the selected term *classification*. In such a task John can be supported by the system through some recommendations suggesting possible relevant parts of the taxonomy where the concept could already exist or where the concept could be inserted.
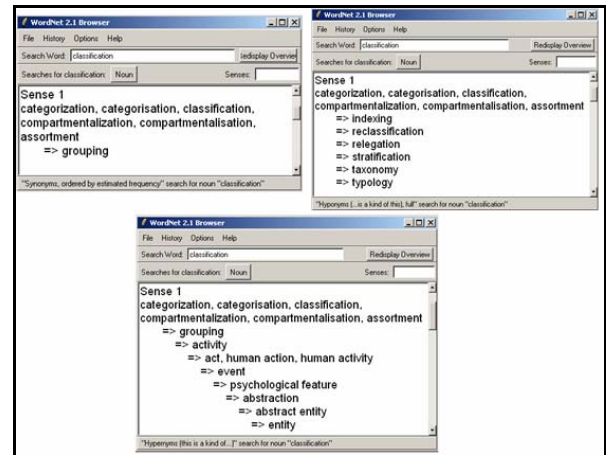

Figure 7. Synonyms, hypernyms and hyponyms for the selected sense of the term *classification*

For example, a possible suggestion can be to attach the new concept as child of *information storage* (Figure 8). If John approves this suggestion a relationship between *information storage* and *classification* will be added and the new taxonomy will be stored in John's personal information space. From now on, the digital library will keep track of new concepts in the John's personal taxonomy and additions of new concepts will be checked to avoid inconsistencies. The selected citation will be automatically classified in John's personal space, according to the new concept just added (Figure 9).

While browsing the space of metadata, John can select and agree with another term, such as *collaborative tagging* which could not have any associated sense in Wordnet. In this case John has not to disambiguate any sense but he has to provide a brief description of the concept. Anyway John has to find the right place in the taxonomy where to insert the concept corresponding to the selected term.

John can also disagree with a term in the space of metadata, in such a situation he can optionally propose new terms. Proposing a new term renders the same scenario as if he has chosen an existing one in the space of metadata.
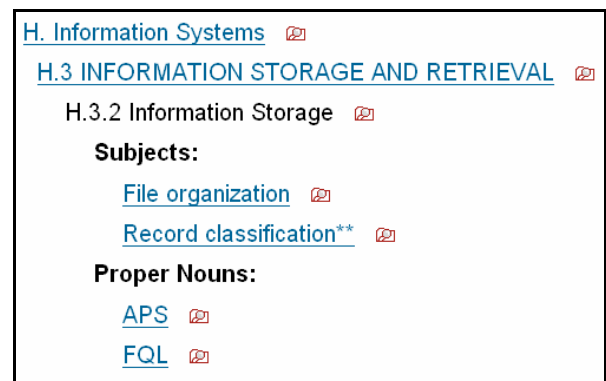

Figure 8. Suggested taxonomy branch where to attach the concept associated with the term *classification*
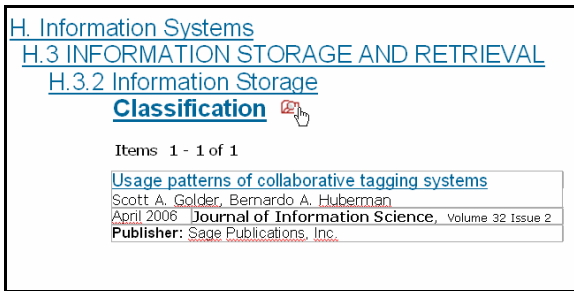
Figure 9. Personal taxonomy

3.2.3 Sharing

John's information space will be structured in a set of binders where he will store citations classified according to his favorite metadata. Moreover, storing and annotating citations will give rise to an evolving personal taxonomy which John can exploit to browse through his personal space. Using the digital library, a user profile will be created in order to keep track of topics of interest. For each binder created by John, one or more corresponding topics of interest will be included in his profile (Figure 10).
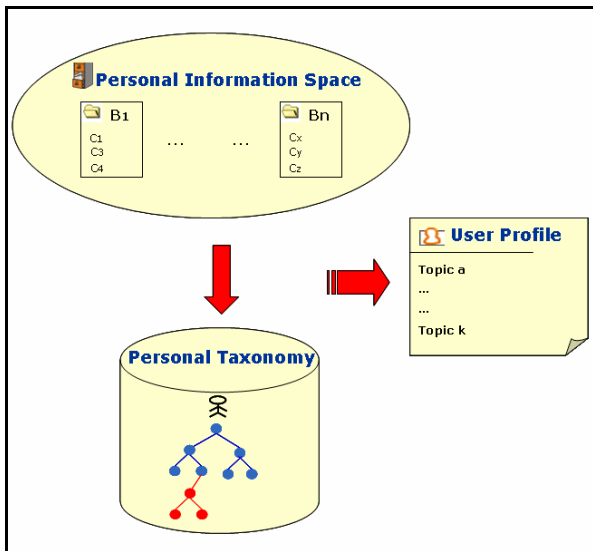

Figure 10. Creation of a user profile

John now chooses to share the binder just created, named *tagging patterns*. Within John's profile the systems looks for one or more topics of interest associated to that binder. Having established the topic of the shared binder, the system looks for other profiles with the same topic, in order to find users which share similar interests with John.

For example two other users, Michael and Lucia have in their profiles analogous topics about collaborative tagging dynamics. Michael has in his personal space a shared binder named *tagging studies*, with the same citation stored by John and other two citations, respectively named "Tagging, communities, vocabulary, evolution" and the other titled "HT06, tagging paper, taxonomy, Flickr, academic article, to read". Figure 11 shows a portion of the Michael's personal taxonomy which describes how Michael has classified citations within his shared binder.
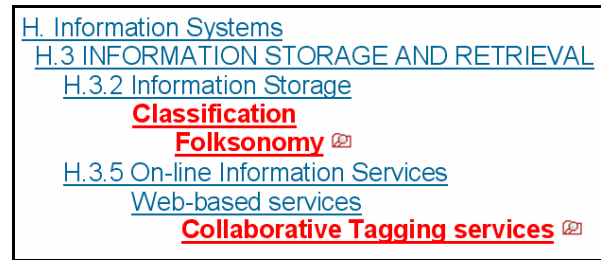

Figure 11. A portion of Michael's personal taxonomy

Lucia has shared a binder named *tagging systems analyses* where she stored all the citations in the Michael's binder and the citation named "What goes around comes around: an analysis of del.icio.us as social space". In Figure 12 there is the portion of Lucia's personal taxonomy relative to all the citations in her shared binder.
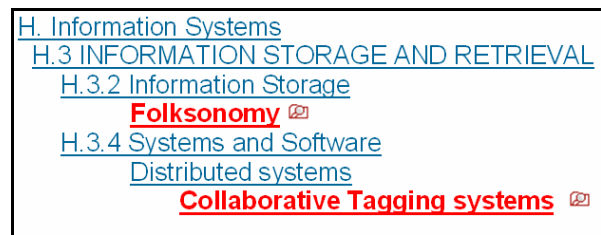

Figure 12. A portion of Lucia's personal taxonomy

Once John has shared the binder, he gains access to a shared information space concerning a particular topic related to the binder. In this shared space, John can view all users interested in the same topic, all citations relevant to the topic stored by these users, as well as one or more shared taxonomies. Every taxonomy in this shared space has the purpose to represent a particular perspective on that topic, depicting a common way to classify related citations employed by a group of people with similar interests. One or more shared portions of these taxonomies are recommended to John. He is now allowed to rank suggestions in accordance with his own perspective. As a result, the shared information space will be displayed to John (Figure 13).

Now John can perform any of the following actions:

- browse through users' personal information spaces, viewing user profiles, taxonomies, shared binders, unless they have been kept as private;
- discover new citations about the topic *collaborative tagging* and add them to either the shared binder or a new one;
- observe how shared taxonomies have been ranked by other users and express his own grade.

After John has shared his binder, users, who have previously contributed to the shared space, will be notified about changes. Afterwards, users can check the information space in order to discover new users with their own similar interests, new citations about the topic, and changes to the shared taxonomies.

John hence contributes to a community perspective for the topic of interest by sharing his personal metadata as well as expressing his preference on the shared taxonomies. On the other hand, he gets feedback for his personal organization while actively taking part to the community.

Topic: *Collaborative tagging dynamics*

Users	Citations	Taxonomies

☃ John
 	
☃ Michael
 	
☃ Lucia
	📄 *Tagging, communities, vocabulary, evolution*
	
	📄 *HT06, tagging paper, taxonomy, Flickr, academic article, to read*
	
	📄 *What goes around comes around: an analysis of del.icio.us as social space*
	
	📄 *Usage patterns of collaborative tagging systems*
		H. Information Systems
		  H.3 INFORMATION STORAGE AND RETRIEVAL
		    H.3.2 Information Storage
		      Classification
		        Folksonomy
		⭐⭐☆☆☆ 2 votes
		
		H. Information Systems
		  H.3 INFORMATION STORAGE AND RETRIEVAL
		    H.3.5 On-line Information Services
		      Web-based services
		        Collaborative Tagging services
		⭐☆☆☆☆ 1 votes
		
		H. Information Systems
		  H.3 INFORMATION STORAGE AND RETRIEVAL
		    H.3.4 Systems and Software
		      Distributed systems
		        Collaborative Tagging systems
		☆☆☆☆☆ 0 votes

Figure 13. The resulting shared information space

4. RELATED WORK

While our approach aims to apply collaborative tagging concepts to the problem of knowledge evolution, much research work assumes the opposite perspective: Discovering semantic relations among tags to enhance how current collaborative tagging systems work.

Mika [11] extends the traditional bipartite model of ontologies with the social dimension leading to a tripartite model of ontologies with three different classes of nodes, namely persons, concepts, and instances and hyperedges representing the commitment of a person in terms of classifying an instance as belonging to a certain concept. This model is exploited by generating two kinds of association networks: the network of concepts and instances and the network of people and concepts. From the association network of concepts and instances, it is extracted a classification hierarchy. From the network of people and concepts, the author generates a hierarchy based on sub-community relationships.

Hotho et al. [9] propose an adaptation of a data mining approach to detect emergent semantics within a collaborative tagging system. The adaptation lies in reducing the three-dimensional folksonomy to a two-dimensional formal context in order to apply association rule mining techniques. Discovered association rules can be then exploited in a recommender system which supports the user in choosing useful tags. The obtained rules can be also seen as subsumption relations, in order to learn a taxonomic structure.

In [8] authors present an algorithm that tries to address the basic level variation issue by converting a large corpus of tags into a navigable hierarchical taxonomy. Tags are grouped using vectors according to the number of times each tag has been used for every annotated resource. Then, the algorithm defines a function to calculate similarity between vectors and a threshold to prune irrelevant values. Finally, for a given dataset a tag similarity graph is created exploiting the social network notion of graph centrality. Starting from the similarity graph and according to three fundamental hypotheses, namely hierarchy representation, noise and general-general assumptions, a latent hierarchical taxonomy is extracted.

Wu et al. [18] exploit a probabilistic generative model to represent the user's annotation behavior in a social bookmarking system and to automatically derive the emergent semantics of the tags. Starting from the assumption that tags heavily used by users with similar interests are semantically related, the authors apply statistical techniques to discover semantic relationships from the different frequencies of co-occurrences among users, resources and tags. The resulting emergent semantics of user interests, tags and web resources is then exploited to develop an intelligent semantic search system with the purpose to search and discover semantically-related web resources.

5. CONCLUSION

This paper provides a community-driven approach to knowledge evolution. Although we have depicted scenarios for a research community, the proposal applies to other online communities.

As in collaborative tagging systems, the main idea is to shift the creation of metadata from a restricted to a collective activity, but still maintaining the expressiveness an ontology can provide for classification.

Knowledge engineers struggle to capture all the variety taking place within a lively community. We hypothesize that augmenting users' participation in the process of annotating and classifying shared items reflects the community knowledge more effectively than relying on prescribed knowledge structures, maintained by a central authority. A collaborative approach to knowledge evolution can split costs over a wide group of people, who have special interests in specific knowledge domains.

The scenarios presented in this paper point out how challenging is to directly involve users in the knowledge evolution process. We need to provide tool support to allow community members to easily organize their personal information spaces, and contribute with a minimal overload. We intend to develop a software agent which is able to monitor users' interactions with the system and learn about users' interests. The agent will gain access to metadata in users' personal information spaces to discover topics of interest. In order to enable software agents to better handle metadata, users' tags will be rendered as RDF statements rather than simple keywords expressed in natural language.

The approach presented here is a first step toward a collaborative knowledge evolution system with the aim to provide an enhanced infrastructure supporting the ever-evolving community knowledge through the active participation of its members.

6. REFERENCES

[1] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001).

[2] Davies, J., Fensel, D., and Harmelen, F. Towards the Semantic Web: Ontology-driven Knowledge Management. John Wiley & Sons, 2003.

[3] Fowler, M. Uml Distilled: A Brief Guide to the Standard Object Modeling Language. Addison-Wesley Professional, 2004.

[4] Furnas, G., Landauer, T., Gomez, L., and Dumais, S. The vocabulary problem in human-system communication, Communications of the ACM, 30, 11 (1987), 964-971.

[5] Golder, S. and Huberman, B. Usage patterns of collaborative tagging systems, Journal of Information Science, 32, 2 (2006), 198-208.

[6] Gruber, T.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal Human-Computer Studies 43 (1993), 907-928.

[7] Haase, P., Völker, J., and Sure, Y. Management of dynamic knowledge, Journal of Knowledge Management, 9, 5 (2005), 97-107.

[8] Heymann, P., Garcia-Molina, H.: Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Technical Report InfoLab 2006-10, Department of Computer Science, Stanford University, Stanford, CA, USA (2006).

[9] Hotho, A., Jäschke, R., Schmitz, C., Stumme, G. Emergent Semantics in BibSonomy. Proc. Workshop on Applications of Semantic Technologies, Informatik 2006, Dresden, 2006.

[10] Mathes, A. Folksonomies-Cooperative Classification and Communication Through Shared Metadata. Technical Report, LIS590CMC, Computer Mediated Communication, Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign, 2004.

[11] Mika, P. Ontologies are us: A unified model of social networks and semantics. Proceedings of the 4th International Semantic Web Conference (ISWC 2005), LNCS 3729, Springer-Verlag, 2005.

[12] Mika, P. Social Networks and the Semantic Web: The Next Challenge. IEEE Intelligent Systems 20 (2005).

[13] Shadbolt, N., Berners-Lee, T., and Hall, W. The Semantic Web Revisited, IEEE Intelligent Systems, 21, 3 (2006), 96-101.

[14] Stahl, G. 2000. Collaborative information environments to support knowledge construction by communities. AI Soc. 14, 1 (Apr. 2000), 71-97.

[15] Suthers, D. D. 2005. Collaborative Knowledge Construction through Shared Representations. In Proceedings of the Proceedings of the 38th Annual Hawaii international Conference on System Sciences (Hicss'05).

[16] Thomas, J. C., Kellogg, W. A., Erickson, T. The knowledge management puzzle: Human and social factors in knowledge management. IBM Systems Journal 40, 4 (2001), 863-884.

[17] Vander Wal, T. Folksonomy Definition and Wikipedia.2005.

[18] Wu, X., Zhang, L., Yu, Y.: Exploring social annotations for the semantic web. Proc. of the 15th international conference on World Wide Web (2006), 417-426.

[19] Zhdanova, A. V., Krummenacher, R., Henke, J., and Fensel, D. 2005. Community-Driven Ontology Management: DERI Case Study. In Proceedings of the the 2005 IEEE/WIC/ACM international Conference on Web intelligence (Wi 2005), IEEE Computer Society Press.