Community-Driven Ontology Evolution Based on Folksonomies

Domenico Gendarmi and Filippo Lanubile

University of Bari, Dipartimento di Informatica, Via E. Orabona, 4 – 70125, Bari, Italy
{gendarmi,lanubile}@di.uniba.it

Abstract. The Semantic Web mission is to enable a better organization of the Web content to improve the searching, navigation and integration of the available information. Although the Semantic Web is intended for machines, the process of creating and maintaining it is a social one: only people, for example, have necessary skills to create and maintain ontologies. While most existing ontologies are designed by single individuals or small groups of experts, actual ontology users are not involved in the development process. Such an individual approach in creating ontologies, lead to a weak community grounding. On the other hand, Social Software is becoming increasingly popular among web users, giving opportunities to exploit the potential of collaboration within a community. Tools like wikis and folksonomies allow users to easily create new content and share contributions over a social network. Social Software tools can go beyond their current limits, by exploiting the power provided by semantic technologies. Conversely, Semantic Web tools can benefit from the ability of Social Software in fostering collaboration among users, by lowering entry barriers. In this paper we propose a new approach for ontology evolution, considering collaborative tagging systems as an opportunity to complement classic approaches used in maintaining ontologies.

Keywords: folksonomy**,** collaborative tagging, ontology evolution, social software, semantic web, semantic collaboration.

1 Introduction

The Semantic Web mission is to enable a better organization of the Web content to improve the searching, navigation and integration of the available information [2]. Current Semantic Web tools require a significant expertise level from their users, specifically in languages and techniques for knowledge representation. Other than being thought for computer science graduates, these tools do not provide any support for collaborative work.

Since Tim Berners-Lee's vision, online communities have taken an active role in the task of knowledge contribution on the Web. The recent phenomenon of Web 2.0 [13], also known as Social Software, has led to the growth of several tools which have succeeded in making this task more attractive to a broader audience. Social Software users are more than just passive information consumers but active participants, working in close collaboration to create new content and share it on the Web.

R. Meersman, Z. Tari, P. Herrero et al. (Eds.): OTM Workshops 2006, LNCS 4277, pp. 181 – 188, 2006.

There are interesting research challenges at the intersection of Social Software and Semantic Web. Social Software can overcome its current limitations, by exploiting the power provided by semantic technologies in searching, navigation and integration of the information published on the Web. Semantic Web can benefit from the ability of Social Software in fostering collaboration among users, then lowering entry barriers to knowledge management.

In this paper we begin to investigate these opportunities, by presenting an approach which aims to combine the main benefits resulting from a specific type of social software, folksonomies, and well-defined formalisms applied in classical knowledge engineering techniques.

In the next section we discuss the drawbacks in knowledge creation and sharing. Section 3 introduces background information related to folksonomies, while Section 4 provides a short overview of known approaches to combine folksonomies and semantic techniques. Finally, Section 5 presents our approach and draws research directions.

2 Issues in Knowledge Sharing

Ontologies play a central role in the Semantic Web vision because they establish common vocabularies and semantic interpretations of terms accessible by machines. Sharing a common understanding of the structure of information among people and software agents is one of the main reasons for using ontologies [7].

Although the Semantic Web is intended for machines, the process of creating and maintaining knowledge is human-intensive. While most existing ontologies are designed by individuals or small groups of experts, ontology users are not involved in the development process. Such a restrictive approach in creating ontologies, leads to a weak community grounding.

The achievement of a widespread participation and shared consensus among ontology users is hampered by entry barriers, like the lack of easy-to-use and intuitive tools capable to include users in the ontology development process. The Ontology Web Language (OWL), for example, has the major drawback to be in several aspects non-intuitive for people who are not familiar with the Description Logics field.

Another relevant problem is the temporal extent of reliable knowledge which tends to be short. More information users learn, more the agreement and consensus among them evolve; thus new pieces of knowledge have to be committed and older pieces have to be constantly checked and validated. However, current ontologies require that all the changes have to be captured and introduced by the same knowledge engineers who created them. The delay in realizing and introducing changes can take weeks or even months, a period of time unacceptable in many dynamic domains, where knowledge regularly and rapidly changes. To be really effective, ontologies thus need to change as fast as the parts of the world they describe [8].

3 Collaborative Tagging

The term folksonomy, a combination of the words “folk” and “taxonomy”, is meant to indicate the act of collaboratively tag resources within Internet communities [16].

Personal user's metadata, commonly called tags, can be applied to any public resource and shared among all the community participants. A folksonomy, also known as collaborative tagging system, exploits the classification activity performed by each user, creating a network of users, resources and tags with a flat structure and no limits in evolution.

People have been starting to legally publish and share their content on the web, in the form of bookmarks, academic paper references, pictures, short audio and video files. Services like del.icio.us, CiteULike, Flickr, Last.fm and YouTube empower users to organize all these information resources through simple keywords.

Keyword-based systems for organizing digital content have already existed for many years. Digital libraries use metadata to discriminate between relevant and irrelevant information needs. While metadata have been long-used to improve information organization and discovery, they are usually created by knowledge engineers or by the original authors [10].

The key element underlying the proliferation of collaborative tagging system is the opportunity for users to collaborate in categorizing all the available resources with no forced restrictions on the allowed terms. In addition to individual benefits coming from the lack of vocabulary restrictions, the whole community can achieve significant advantages resulting from each participant contribution.

Furthermore, collaborative tagging systems create a strong sense of community among their users. Users are able 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. Marginal opinions can coexist with popular ones without disrupting the implicit emerging consensus on the meaning of the terms rising up the folksonomy.

Opposed to professionally developed taxonomies also called controlled vocabularies, folksonomies show interesting potential to overcome the vocabulary problem [4]. 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 [11]. While professionally created metadata are characterized by high quality, they do not scale up. On the contrary, collaborative tagging systems have a very short learning curve because there is not a predefined structure and syntax to learn.

The most relevant strength of a folksonomy is its ability in adhering to the personal way of thinking. There is no need for establishing a common agreement on the meaning of a tag because it gradually emerges with the use of the system. Folksonomies can then react quickly to changes and be responsive to new user needs.

4 Hybrid Folksonomy-Based Approaches

Folksonomies and ontologies have been often seen as competitors, the hype generated by collaborative tagging has inspired long debates over the web about which can be the best approach to categorization [12], [15]. Recently the debate has moved towards scholarly literature, analysing collaborative tagging phenomenon more rigorously.

Golder and Huberman [5] analyse data gathered from a popular folksonomy, del.icio.us, to better understand the structure of tagging systems, as well as their dynamic aspects. The study examines how tags are used over time by users and how they eventually stabilize, for a specific resource, over time. Findings show a high variety in the sets of tags employed by users. The study also discovers some measure of regularity in user activity, tag frequencies, kinds of tags used and bursts of popularity in bookmarking. On this basis, a dynamic model is proposed to predict stable tagging patterns. Golder and Huberman also discuss difficulties of folksonomies related to semantic and cognitive aspects of tags, such as problems in synonymy, polysemy and basic level variation.

To overcome common issues in using tags, some new approaches, bent to exploit semantic techniques in collaborative tagging, have recently appeared in literature.

Heymann and Garcia-Molina [9] describe an algorithm which aims to address the basic level problem by converting a large corpus of tags into a navigable hierarchical taxonomy. Tags are aggregated into vectors denoting the number of times a tag has been used for each annotated resource. A similarity function is defined and calculated using the cosine similarity between vectors, then a threshold is established 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 mined. The algorithm is tested on two different datasets, from del.icio.us and CiteULike, but only in the first case results are promising.

In [17] the authors propose statistical techniques to mine the implicit semantics embedded in the different frequencies of co-occurrences among users, resources and tags in folksonomies. A probabilistic generative model is used to represent the user's annotation behaviour in del.icio.us and to automatically derive the emergent semantics of the tags. Synonymous tags are grouped together and polysemous tags are identified and separated. Moreover, the derived emergent semantics is exploited to discover and search shared web bookmarks. The initial evaluation shows that such a method can effectively discover semantically related bookmarks.

Similarly to the previous work but with different aims, in [1] clustering techniques are applied to folksonomies. The authors propose a more effective searching, subscribing and exploring service which automatically generate clusters of tags. For tags used for the same resource, the algorithm counts the number of co-occurrences of any pair of tags. A cut-off point is established to distinguish between strongly and weakly related tags which are represented in an undirected weighted graph. The authors exploit an algorithm for graph clustering which introduces the modularity function for measuring the quality of a particular partition of nodes in a graph. The algorithm is tested on the RawSugar database and results are available at the RawSugar lab page[1].

Xu et al. [18] define a set of general criteria for a good tagging system to identify the most appropriate tags, while eliminating noise and spam. These criteria, identified through a study of tag usage by real users in My Web 2.0, cover desirable properties of a good tagging system, including high coverage of multiple facets to ensure good

[1] http://www.rawsugar.com/lab

recall, least effort to reduce the cost involved in browsing, and high popularity to ensure tag quality. The authors propose a collaborative tag suggestion algorithm using previous criteria to recommend high-quality tags. The proposed algorithm employs a goodness measure for tags derived from collective user authorities to contrast spam. The goodness measure is iteratively adjusted by a reward-penalty algorithm, which incorporates content-based and auto-generated tags. These principles ensure that the suggested tag combination has a good balance between coverage and popularity. Preliminary results, coming from an experiment conducted on My Web 2.0, show that such an algorithm is effective in suggesting appropriate tags that match the expected properties.

Schmitz [14] explores a model that leverages statistical natural language processing techniques to induce ontology from Flickr database. An existing probabilistic subsumption based model is adapted to existing tags set, adjusting the statistical thresholds to reflect the ad hoc usage, and adding filters to control the highly idiosyncratic Flickr vocabulary. The proposed model produces subtrees that generally reflect distinct facets, but can not categorize concepts into facets. Although resulting trees are manually evaluated, early results are promising compared to related subsumption models. Moreover a series of refinements to the model are planned to improve the accuracy and induce a faceted ontology.

Finally, Gruber's proposal [6] is the establishing of a common format to achieve the interoperability among different tagging applications. Tags are considered as a form of voting and the act of tagging performed by users is compared to the innovation of incorporating the hyperlink as a measure of popular acclaim pioneered by Google. Two use cases, illustrating current issues in sharing tags across multiple applications, are presented as motivations of the need of a common conceptualization for representing tags meaning. The author proposes to create an ontology for tagging, the “TagOntology”, meant to identify and formalize a shared conceptualization of the tagging activity, as well as to develop the technology that commits to the ontology at the semantic level.

5 Towards a Collaborative Approach for Ontology Evolution

While most efforts, such as the above mentioned works, are focused on how to create better folksonomies, we propose an approach going to the reverse direction. Our goal is to exploit what is already provided by existing ontologies, such as explicit semantics and support for reasoning in combination with the ability of folksonomies to foster collaboration within a community.

The vision we propose is a community of autonomous and networked users who cooperate in a dynamic and open environment. Each participant will organize some piece of knowledge according to a self-established vocabulary, create connections and negotiate meaning with other users within the community.

Augmenting the involvement of users, by enabling community members to actively participate to the ontology evolution process is a key factor to achieve a community common ground. Starting from an existing ontology and allowing users to freely edit ontology classes, according to their personal vocabulary, can significantly improve the ontology maintenance process, complying with the knowledge drift.

However, current ontologies are usually manually written in standard formal languages, such as OWL[2], through standalone toolkits, such as Protégé[3]. The requirement of background skills, such as the knowledge of the OWL syntax, and the lack of tools enabling a distributed and collaborative contribution to the ontology enrichment can severely hinder our vision.

Given an initial ontology, we propose a collaborative ontology evolution system, which allows community members to add, modify, or delete existing and new ontology classes, according to their own needs. The editing of the classes does not require any special skill but it is allowed through the use of simple metadata, like adding tags in current tagging systems.

An open rating system will also be provided, each time a user contributes to the evolution of the ontology this can be seen as a voting activity. Adding or editing a class can be seen as a vote for the new class added, while contributing to the ontology without editing existing classes (e.g. inserting new individuals) can be considered as a vote for an existing class. A measure of the quality of an ontology class can be thus calculated according to the weight average of all the votes obtained by the class. This rating system can lead the most popular names for a specific ontology class to rise to the top, while the less exploited ones will fall to the bottom, similarly to the behaviour of a tag cloud in a folksonomy.

On the other hand, to support this collaborative ontology evolution, an environment providing distributed access and supporting itself group activity, such as, simultaneous viewing and editing, is needed. While distributed access can be provided by a common web interface the collaborative work can be ensured by the wiki technology.

Wikis are themselves collaborative tools providing a community with web writing and browsing functionalities. Wiki systems typically provide an easy-to-use editing environment to create or modify content on the fly requiring no tool but the browser. They usually have a version control system to record modification of contents in order to show-up recent changes and the versions history, a user profile and concurrent conflict management system to enable multiple user editing the same contents, and a content navigation system that simplifies indexing, searching and linking wiki pages within a wiki system.

On this basis, our purpose is to use the wiki technology to edit ontologies via the web, thus developing a Web Ontology Editor. A web-based interface will provide features to support collaborative editing, ontology evolution tracking (e.g., identify classes that have been added, deleted or modified), as well as browsing functionalities that allow users to search and navigate the ontology.

An ontology module can be composed of one or more wiki pages, multiple users can edit the same content with version control and transaction management, and the ontology can be managed like a common wiki repository. Moreover, most wikis offer the opportunity to define customized markup. Users can associate predefined sequences of characters with commands that the wiki engine can interpret and execute, in order to render a new sequence of characters expressed in a different syntax. Therefore it is possible to define a set of custom markup tags corresponding to

---

[2] http://www.w3.org/TR/owl-features/

[3] http://protege.stanford.edu/

the syntax of the ontologies, such as OWL or RDF syntax. When a wiki page is under editing, its custom markup is translated to user friendly text, such as HTML web page. This customized Wiki markup can be expressed in human readable syntax and like in current wikis users could create new contents simply by adding and modifying the source of an existing one.

Adopting a collaborative approach for ontology maintenance is a challenging research topic for the benefits it can bring to conventional approaches [3].

Ontologies which are improved and used as a community reflect the knowledge of users more effectively than ontologies maintained by knowledge engineers who struggle to capture all the variety taking place within a lively community.

Collaborative ontology editors can strengthen community participation in ontology development and maintenance process because users are enabled to autonomously change knowledge and look at changes that are triggered by their actions.

Furthermore, classic ontology maintenance is expensive as one or more knowledge engineers have to be called on purpose. With community-driven ontology evolution, costs are split over a wide group of people who have a special interest in maintaining the ontology they use up-to-date.

The proposed approach can be a first step toward a collaborative system capable of allowing ontologies to evolve mainly through the contribution of its users. A web ontology editor can relieve users of the system to know OWL syntax and allow them to contribute to the ontology by adding tags as proxies of metadata.

Future work plan includes the development and evaluation of the proposed web ontology editor. We aim to show that such an environment can scale up and support collaboration among several users. A further step will be to explore how semantics can be achieved through collaboration among users without burdening the user experience.

References

1. Begelman, G., Keller, P., Smadja, F.: Automated Tag Clustering: Improving search and exploration in the tag space. Proc. Of WWW2006, Collaborative Web Tagging Workshop (2006)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001)
3. Davies, J., Fensel, D., van Harmelen, F.: Towards the Semantic Web: Ontology-driven Knowledge Management. John Wiley & Sons, Inc., New York, NY (2003)
4. Furnas, G. W., Landauer, T. K., Gomez, L. M., Dumais, S. T.: The vocabulary problem in human-system communication. Communications of the ACM 30, 11 (1987) 964-971
5. Golder, S., Huberman, B.: Usage patterns of collaborative tagging systems. Journal of Information Science 32, 2 (2006) 198-208
6. Gruber., T.: Folksonomy of Ontology: A Mash-up of Apples and Oranges. First on-Line conference on Metadata and Semantics Research (MTSR'05) (2005)
7. Gruber., T.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal Human-Computer Studies 43 (1993) 907-928
8. Haase, P., Volker, J., Sure, Y.: Management of dynamic knowledge. Journal of Knowledge Management 9, 5 (2005) 97-107

9. 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)
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 (2004)
11. McCulloch, E., Macgregor, G.: Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. Library Review 55, 5 (2006)
12. Merholz, P.: Clay Shirky's Viewpoints are Overrated. (2005)
13. O'Reilly, T.: What is Web 2.0. (2005)
14. Schmitz, P.: Inducing Ontology from Flickr Tags. Proc. Of WWW2006, Collaborative Web Tagging Workshop (2006)
15. Shirky, C.: Ontology is Overrated: Categories, Links, and Tags. (2005)
16. Vander Wal, T.: Folksonomy Definition and Wikipedia. (2005)
17. 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
18. Xu, Z., Fu, Y., Mao, J., Su, D.: Towards the Semantic Web: Collaborative Tag Suggestions. Proc. Of WWW2006, Collaborative Web Tagging Workshop (2006)