

Chapter 46

Representing and Sharing Tagging Data Using the Social Semantic Cloud of Tags

Hak-Lae Kim

National University of Ireland, Galway, Ireland

John G. Breslin

National University of Ireland, Galway, Ireland

Stefan Decker

National University of Ireland, Galway, Ireland

Hong-Gee Kim

Seoul National University, South Korea

ABSTRACT

Social tagging has become an essential element for Web 2.0 and the emerging Semantic Web applications. With the rise of Web 2.0, websites that provide content creation and sharing features have become extremely popular. These sites allow users to categorize and browse content using tags (i.e., free-text keyword topics). However, the tagging structures or folksonomies created by users and communities are often interlocked with a particular site and cannot be reused in a different system or by a different client. This chapter presents a model for expressing the structure, features, and relations among tags in different Web 2.0 sites. The model, termed the social semantic cloud of tags (SCOT), allows for the exchange of semantic tag metadata and reuse of tags in various social software applications.

INTRODUCTION

With the rise of Web 2.0, websites which provide content creation and sharing features have become extremely popular. Many users have become actively involved in adding specific metadata in the

form of tags and content annotations in various social software applications. While the initial purpose of tagging is to help users organize and manage their own resources, collective tagging of common resources can be used to organize information via informal distributed classification systems called *folksonomies* (Mathes, 2004; Merholz, 2004).

DOI: 10.4018/978-1-60566-368-5.ch046

Studies of tagging and folksonomies can be divided into two main approaches: (a) semantic tagging concentrates on folksonomies that are inconsistent and even inaccurate because a large group of untrained users assign free-form terms to resources without guidance. Since this approach aims to resolve tag ambiguities, a wealth of ideas and efforts is emerging regarding how to use and combine ontologies with folksonomies (Weller, 2007); (b) social networking focuses on a community of users interested in a specific topic that may emerge over time because of their use of tags (Mika, 2005). The power of social tagging lies in the aggregation of information (Quintarelli, 2005). Aggregation of information involves social reinforcement by reinforcing social connections and providing social search mechanisms. Thus, a community built around tagging activities can be considered a social network with an insight into relations between topics and users.

Using freely determined vocabularies by a participant is less costly than employing an expert (Sinclair & Cardew-Hall, 2007) and a cognitive load of tagging in comparison with taxonomies or ontology is relatively low (Merholz, 2004). However, tagging the data from social media sites without a social exchange is regarded as an individual set of metadata rather than a social one. Although tagging captures individual conceptual associations, the tagging system itself does not promote a social transmission that unites both creators and consumers. To create social transmission environments for tagging, one needs a consistent way of exchanging and sharing tagging data across various applications or sources. In this sense, a formal conceptual model to represent tagging data plays a critical role in encouraging its exchange and interoperation. Semantic Web techniques and approaches help social tagging systems to eliminate tagging ambiguities.

BACKGROUND

Social Tagging

Social tagging and folksonomies have received much attention from the Semantic Web and Web 2.0 communities as a new way of information categorization and indexing. Among the most popular websites that employ folksonomies are Del.icio.us¹ (social bookmarking system) and Flickr² (photo-sharing network). CiteULike, using a similar approach to Del.icio.us, focuses on academic articles.³ There are a number of multimedia sites that support tagging, such as Last.fm⁴ for music and YouTube⁵ for video.

Although the idea of a *tag* is not new, most people agree that a tag is no longer just a keyword. There is semantic information associated with a tag (Weller, 2007). A tag represents a type of metadata used for items such as resources, links, web pages, pictures, blog posts, and so on. Tagging can be defined as a way of representing concepts through keywords and cognitive association techniques without enforcing a categorization. The term *folksonomy* is a fusion of the two words *folk* and *taxonomy* (Vander Wal, 2004); it became especially popular with the proliferation of web-based social software applications, such as social bookmarking or annotating photographs. Building on the above definitions, folksonomy can be considered as a collaborative practice and method of creating and managing tags for the purpose of annotating and categorizing content (Mathes, 2004).

Advantages and disadvantages of social tagging present an issue for discussion. Although social tagging and folksonomies have much to offer users who utilize tags in various social media sites, there are important drawbacks inherent within the current tagging systems: for example, there is no formal conceptualization to represent tagging data in a consistent way and no interoperability support for exchanging tagging data among different applications or people (Marlow et al., 2006; Kim et al., 2007). The simplicity and accessibility of

tags may lead to a lack of precision resulting in keyword ambiguity caused by misspelling certain words, as well as using synonyms, morphologies, or over-personalized tags. Since there are many different ways of using tags, it may be easy to misunderstand the meaning of a given tag.

Aside from these problems, social tagging systems do not provide a uniform way to share and reuse tagging data amongst users or communities. There is no consistent method for transferring tags between the desktop and the web or for reusing one's personal set of tags between either web-based systems or desktop applications. Although some folksonomy systems support an export functionality using their Open APIs (Application Programming Interfaces) and share their data with a closed agreement among sites, these systems do not offer a uniform and consistent way to share, exchange, and reuse tagging data for leveraging social interoperability. Therefore, it may be difficult to meaningfully search, compare, or merge similar tagging data from different applications.

With the use of tagging systems increasing daily, these limitations will become critical. The limitations come from lack of standards for tag structure and the semantics for specifying the exact meaning. To overcome the current limitations of tagging systems, it may be beneficial to take into account not only standards for representing tagging data but also develop interoperable methods to support tag sharing across heterogeneous applications.

SEMANTIC TAGGING APPROACHES

Folksonomy vs. Ontology

In general, a *taxonomy* is the organization of a set of information for a particular purpose in a hierarchical structure. An *ontology* is set of well-defined concepts describing a specific domain. It has strict and formal rules for describing relationships among concepts and for defining properties. The distinction between an ontology and a

taxonomy is sometimes vague. A simple ontology without properties and constraints (i.e., concept hierarchy) could be considered a taxonomy, but a heavyweight ontology should clearly specify its capabilities.

From a classification perspective, folksonomies and ontologies can be placed at the two opposite ends of the spectrum. When compared to a traditional classification system, a folksonomy can be seen as a set of terms forming part of a flat namespace; that is, a folksonomy is a completely uncontrolled and flat system (Tonkin, 2006). To its disadvantage, folksonomy has no hierarchy and there are no directly specified parent-child relationships between the varying descriptions of the same object. Despite these limitations, the usefulness of folksonomies has been acknowledged: a folksonomy is a user-generated classification created through a bottom-up consensus.

Tag Ontology

There are certain disagreements on the merits of folksonomies and traditional classifications (see, for example, Hendler, 2007; O'Reilly, 2005; Shirky, 2004; Spivack, 2005). Shirky (2004) makes an argument that ontological classification or categorization is overrated in terms of its value. Shirky views folksonomies as emergent patterns in users' collective intelligence and claims that they can be harnessed to create a bottom-up consensus view of the world. According to Shirky, traditional classification systems have been structured using hierarchical taxonomies by experts studying a particular domain. Therefore these systems do not satisfy user-specific ways of thinking and organizing information. Meanwhile, Gruber (2005) criticizes Shirky's approach in that he fails to point out that a folksonomy has limitations to represent, share, exchange, and reuse tags and confuses "ontology-as-specified-conceptualization" with a very narrow form of specification. Hendler (2007) also argues that Shirky misunderstood how ontologies could be built on the principles

of the Semantic Web. Spivack (2005) asserts that folksonomies are just specific, highly simplistic cases of ontologies with minimal semantics.

Despite conflicting differences between folksonomies and ontologies, the Semantic Web and ontologies can be seen as a complement to folksonomies. Gruber (2005) and Spivack (2005) emphasize the importance of folksonomies and ontologies working together. In particular, Gruber (2005) proposes the “Tag Ontology.” This aims at identifying and formalizing a conceptualization of the activity of tagging, and building technology that commits to the ontology at a semantic level. This approach is a good starting point to bridge Web 2.0 and the Semantic Web:

- **Gruber’s Conceptual Model:** Suggests a model that defines a tagging activity including an object, a tag, a tagging, and a source.
- **Richard Newman’s Tag Ontology:** Defines the three core concepts of Tagger, Tagging, Tag for representing the tagging activity (Newman, 2005) and is based on a tripartite tagging (i.e., user, resource, and tag).

The two approaches are focused on tagging activities or events that people used to tag in resources using terms. Therefore the core concept is Tagging. The concept of tagging has a relationship, as a concept, with Tagger and Resource to describe people who participate in a tagging event and objects to where a tag is assigned. However, there are no ways to describe the frequency of tags in these ontologies.

SCOT: Social Semantic Cloud of Tags

SCOT (Social Semantic Cloud of Tags) is an ontology for sharing and reusing tag data and representing social relations among individuals. It aims to describe the structure and the semantics

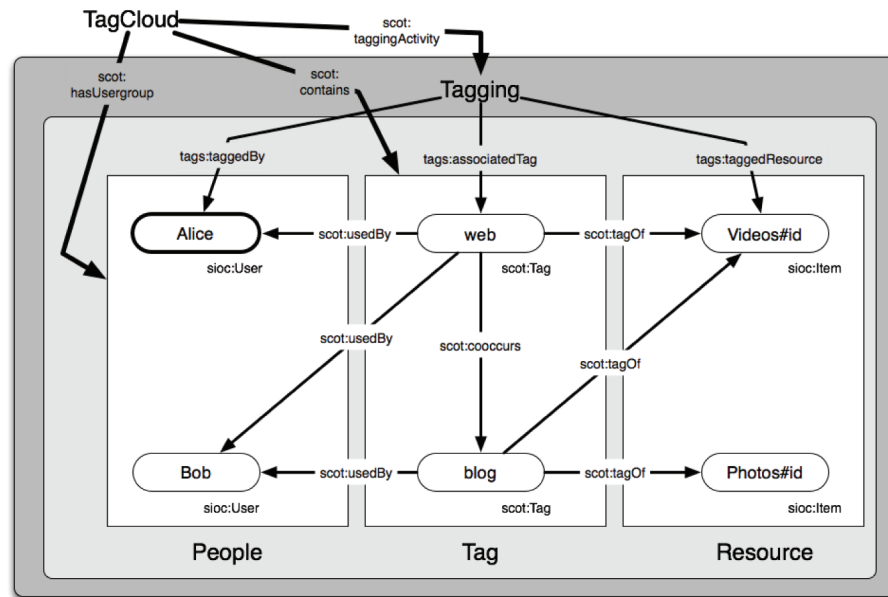
of tagged data as well as offer interoperability of data among different sources.

Figure 1 illustrates the relations among the elements as well as a tagging activity. The vocabularies can be used to make explicit a collection of users, tags, and resources; they are represented by a set of RDF classes and properties that can be used to express the content and structure of the tagging activity as an RDF graph.

The TagCloud, which represents a conceptual model of a folksonomy, has a connection point to combine other concepts (i.e., tag, tagging, tagger, and so on.) and can be connected with other tag cloud with a unique namespace. The `scot:TagCloud` is a subclass of `sioc:Container`. All information consisting of relationships among taggers, items, and all tags is connected to this class. This class has the `scot:contains` and the `scot:hasUsergroup` properties. The former represents tags in a given domain or community while the latter describes taggers who participate in a tagging activity. A tagger may not be a single person according to contexts. For instance, if multiple taggers in a certain community generate a tag cloud, this tag cloud should contain all taggers. The `scot:hasUsergroup` property represents a person with a container. The `scot:composedOf` property describes a part of a TagCloud. In particular, if a TagCloud has more than two tag clouds, the property identifies each tag cloud. The `scot:taggingActivity` property present a relationship between a TagCloud and a Tagging.

The `scot:Tag` class, a subclass of `tags:Tag` (Tag class of Tag Ontology), allows users to assert that a tag is an atomic conceptual resource. A tag is a concept associated with a piece of information. The concept has many different variations according to taggers’ cognitive patterns. Tag ambiguities, one of the most critical problems, result from this reason. The SCOT ontology provides several properties such as `scot:spellingVariant`, `scot:synonym` to solve this problem. It is called the “linguistic property” since these properties focus on representing the meaning of a tag and

Figure 1. Terms and relations in SCOT that can be used to describe tagging activities



the relationships between each tag. In addition, the ontology has properties to describe occurrence of a tag (i.e., `scot:frequency`). A tag itself has its own frequency. The frequency is not unique, but it is an important feature to distinguish or compare with other tags. We called it a “numerical property.” The properties have their own numerical values by computing. The properties in Figure 1 show high-level properties in the SCOT ontology.

In addition to representing the structure and the semantics of tags, the model allows the exchange of semantic tag metadata for reuse in social applications and interoperation amongst data sources, services, or agents in a tag space. These features are a cornerstone to being able to identify, formalize, and interoperate a common conceptualization of tagging activity at a semantic level.

SCOT aims to incorporate and reuse existing vocabularies as much as possible to avoid redundancies and to enable the use of richer metadata descriptions for specific domains. The ontology has a number of properties to represent social tagging activity and relationships among elements occurring in an online community.

- DC, or Dublin Core, provides a basic set of properties and types for annotating documents. In SCOT, we use the properties `dc:title` for the title of a TagCloud, `dc:description` to give a summary of the TagCloud, `dc:publisher` to define what system is generating the TagCloud, `dc:creator` to link to the person who created this set of tags. `dcterms:created`, from the Dublin Core refinements vocabulary, is used to define when a TagCloud was first created.
- FOAF (Brickley & Miller, 2004), or Friend of a Friend, specifies the most important features related to people acting in online communities. The vocabulary allows us to specify properties about people commonly appearing on personal homepages, and to describe links between people who know each other. `foaf:Person` is used to define the creator of a particular TagCloud. `foaf:Group` can be used to define a group of people who have created a group TagCloud.
- SIOC (Breslin et al., 2005), Semantically-

Interlinked Online Communities, provides the main concepts and properties required to describe information from online communities (e.g., message boards, wikis, blogs, etc.) on the Semantic Web. In the context of SCOT, `sioc:Usergroup` can be used to represent a set of `sioc:User` who have created the tags contained within a particular group TagCloud. A TagCloud is also a type of `sioc:Container`, in that it contains a set of Tags (subclass of `sioc:Item`).

- SKOS (Simple Knowledge Organization Systems) provides specifications and standards to support the use of vocabularies, such as thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary as well as terminologies and glossaries (Miles & Bechhofer, 2008). Tag is a subclass of `skos:Concept`, and a number of SKOS properties are used to define the relationships between Tags: broader, narrower, and so on.

Int.ere.st: Platform for Tag Sharing

Int.ere.st is a website where people can manage their tagging data from various sources, search resources based on their tags which were created and used by themselves, and leverage a sharing and exchanging of tagging data among people or various online communities.⁶ The site is a platform for providing structure and semantics to previously unstructured tagging data via various mashups. The tagging data from distributed environments (such as blogs) can be stored in a repository, such as SCOT, via the Mashup Wrapper, which extracts tagging data using Open APIs from host sites. For instance, the site allows users to dump tagging data from Del.icio.us, Flickr, and YouTube; these tagging data are transformed into SCOT instances on a semantic level. Thus, all instances within Int.ere.st include different tagging contexts and connect various people and sources with the same

tags. In addition, users can search people, tags, or resources and can bookmark some resources or integrate different instances. Through this iterative process, the tags reflect distributed human intelligence into the site.

Int.ere.st is the first OpenTagging Platform⁷ of the Semantic Web, since users can manage a collection of tagging data in a smarter and more effective way as well as search, bookmark, and share their own as well as other's tagging data underlying the SCOT ontology. Those functionalities help users exchange and share their tagging data based on the Semantic Web standards. The site is compatible with other Semantic Web applications, and its information can be shared across applications. This means that the site enables users to create Semantic Web data, such as FOAF, SKOS, and SIOC automatically. RDF vocabularies can be interlinked with the URIs of SCOT instances that are generated in the site and shared in online communities.

FUTURE TRENDS

Social computing enables building social systems and software; it also allows for embedding social knowledge in applications rather than merely describing social information. Within social network analysis, traditional approaches have focused on static networks for small groups. As the technologies move forward, a major challenge for social network analysis is to design methods and tools for modeling and analyzing large-scale and dynamic networks. In particular, folksonomies are inherently dynamic and have different contexts among sources.

To facilitate the development of a social network for folksonomies, it is important to pay attention to social information. Although SNA allows analyzing phenomenon of social behavior at both individual and collective level, we do not have a solution that represents the relations among elements and reflects them to the objects. Tag

ontologies are promising in providing the tools and formalism for representing social information including users, resources, tags, and their relationships.

Social search has become an active area in academic research as well as industry. Social search is a type of search engine that determines the relevance of search results by considering user interactions, contributions, or activities, such as bookmarking, tagging, and ranking. For instance, Del.icio.us and Spurl⁸ (social bookmarking services) rely on user rankings, while Technorati⁹ and Bloglines¹⁰ (tag aggregators) analyze blogs and feed-based content. In particular, Swicki¹¹ and Rollyo¹² offer a community-based topic search as well as “searchles”¹³ (these allow users to tag, group, and save links and create their own “SearchlesTV” channels through video mashups).

Most approaches, however, are limited to different types of resources and to show a comprehensive perspective on social relations across different applications. For instance, if users are involved in two different social spaces such as Del.icio.us and YouTube, one cannot build an integrated social network unless the two services have a mutual agreement. This issue, to some degree, is related to social information representation, since both websites have different aspects and events for building social connections. Thus, if one has a common conceptualization for social events, it is easy to build social connections among different spaces and to search them from different sources. Since tag ontologies are suitable to represent common conceptualization of social events, a social search can adopt Semantic Web technologies. We believe that a social search can benefit from a formal conceptualization of social knowledge, including tagging data based on the Semantic Web.

CONCLUSION

Tags have become an essential element for Web 2.0 and the Semantic Web applications. There is a vast collection of user-created content residing on the web. Tagging is a promising technological breakthrough offering new emerging opportunities for sharing and disseminating metadata. The critical issues discussed in this chapter offer many implications and challenges for representing tagging data semantically and exchanging them socially. With emphasis placed on tag ontologies and opportunities, these issues must be confronted without delay. Creators and consumers of folksonomies, as well as service providers, will profit from effective and efficient tagging methods that are socially and semantically enhanced.

REFERENCES

- Breslin, J. G., Decker, S., Harth, A., & Bojars, U. (2005). SIOC: An approach to connect Web-based communities. *International Journal of Web-Based Communities*, 2(2), 133–142.
- Brickley, D., & Miller, L. (2004). *FOAF vocabulary specification*. Retrieved July 15, 2008, from <http://xmlns.com/foaf/0.1>
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198–208. doi:10.1177/0165551506062337
- Gruber, T. (2005). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*, 3(2), 1–11.
- Hendler, J. (2007). *Shirkyng my responsibility*. Retrieved July 15, 2008, from <http://www.mindswap.org/blog/2007/11/21/shirkyng-my-responsibility>

- Kim, H. L., Breslin, J. G., Yang, S. K., & Kim, H. G. (2008). Social semantic cloud of tag: Semantic model for social tagging. In *Proceedings of the 2nd KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications* (pp. 83-92). Berlin, Germany: Springer.
- Kim, H. L., Yang, S. K., Breslin, J. G., & Kim, H. G. (2007). Simple algorithms for representing tag frequencies in the SCOT Exporter. In *Proceedings of Intelligent Agent Technologies* (pp. 536-539).
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. In U. K. Wiil, et al. (Eds.), *Proceedings of the 17th Conference on Hypertext and Hypermedia* (pp. 31-39). New York: ACM Press.
- Mathes, A. (2004). *Folksonomies: Cooperative classification and communication through shared metadata*. Retrieved July 15, 2008, from <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Merholz, P. (2004). *Metadata for the masses*. Retrieved July 15, 2008, from <http://www.adaptivepath.com/ideas/essays/archives/000361.php>
- Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science . Services and Agents on the World Wide Web*, 5(1), 5-15. doi:10.1016/j.websem.2006.11.002
- Miles, A., & Bechhofer, S. (2008). *SKOS Simple knowledge organization system reference*. Retrieved July 15, 2008, from <http://www.w3.org/TR/skos-reference>
- Newman, R. (2005). *Tag ontology design*. Retrieved July 15, 2008, from <http://www.holygoat.co.uk/blog/entry/2005-03-23-2>
- O'Reilly, T. (2005). *What is Web 2.0: Design patterns and business models for the next generation of software*. Retrieved July 15, 2008, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- Quintarelli, E. (2005). *Folksonomies: Power to the people*. Retrieved July 15, 2008, from <http://www.iskoi.org/doc/folksonomies.htm>
- Shirky, C. (2004). *Ontology is overrated: Categories, links, and tags*. Retrieved July 15, 2008, from http://www.shirky.com/writings/ontology_overrated.html
- Sinclair, J., & Cardew-Hall, M. (2008). The folksonomy tag cloud: When is it useful? *Journal of Information Science*, 34(1), 15-29. doi:10.1177/0165551506078083
- Spivak, N. (2005). *Folktologies--beyond the folksonomy vs. ontology distinction*. Retrieved July 15, 2008, from http://novaspivack.typepad.com/nova_spivacks_weblog/2005/01/whats_after_fol.html
- Tonkin, E. (2006). *Folksonomies: The fall and rise of plain-text tagging*. Retrieved July 15, 2008, from www.ariadne.ac.uk/issue47/tonkin/
- Vander Wal, T. (2005). *Explaining and showing broad and narrow folksonomies*. Retrieved July 15, 2008, from <http://www.vanderwal.net/random/category.php?cat=153>
- Weller, K. (2007). Folksonomies and ontologies. 2007. Two new players in indexing and knowledge representation. In H. Jezzard (Ed.), *Applying Web 2.0. Innovation, impact and implementation: Online Information 2007 Conference Proceedings*, London (pp. 108-115). Retrieved July 15, 2008, from http://www.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/35/1204288118weller009_.htm

KEY TERMS AND DEFINITIONS

Folksonomy: A practice and method of collaboratively creating and managing tags for the purpose of annotating and categorizing content. The term *folksonomy* is a fusion of two words: *folk* and *taxonomy*. Folksonomies became popular with the introduction of web-based social software applications, for example, social bookmarking and photograph annotating.

Mashup: Involves web services or applications combining data from different websites. In general, mashup services are implemented by combining various functionalities with open APIs.

Ontology: Is set of well-defined concepts describing a specific domain.

Open API (Application Programming Interface): Is used to describe a set of methods for sharing data in Web 2.0 applications.

Semantic Web: Is an extension of the current World Wide Web that links information and services on the web through meaning and allows people and machines use web content in more intelligent and intuitive ways.

Social Computing: Is defined as any type of collaborative and social applications that offer the gathering, representation, processing, use, and dissemination of distributed social information.

Social Semantic Cloud of Tags (SCOT): Is an ontology for sharing and reusing tagged data and representing social relations among individuals. It aims to describe the structure and the semantics of data and to offer the interoperability of data among different sources.

Social Software: Can be defined as a range of web-based software programs that support group

communication. Many of these programs share similar characteristics, for example, open APIs, customizable service orientation, and the capacity to upload data and media.

Social Tagging: Also known as *collaborative tagging*, refers to assigning specific keywords or *tags* to items and sharing the set of tags between communities of users.

Tag: A type of metadata used for items such as resources, links, web pages, pictures, blog posts, and so on.

Tagging: A way of representing concepts through tags and cognitive association techniques without enforcing a categorization.

Taxonomy: A method of organizing information in a hierarchical structure using a set of vocabulary terms.

ENDNOTES

- ¹ <http://del.icio.us>
- ² <http://www.flickr.com>
- ³ <http://www.citeulike.org>
- ⁴ <http://www.lastfm.com>
- ⁵ <http://www.youtube.com>
- ⁶ <http://int.ere.st>
- ⁷ <http://opentagging.org>
- ⁸ <http://www.spurl.net>
- ⁹ <http://www.technorati.com>
- ¹⁰ <http://www.bloglines.com>
- ¹¹ <http://www.swicki.com>
- ¹² <http://rollyo.com>
- ¹³ <http://www.searchles.com>