Chapter 11. Towards OpenTagging Platform using Semantic Web Technologies

Hak Lae Kim

DERI, National University of Ireland, Galway, Ireland

John G. Breslin

DERI, National University of Ireland, Galway, Ireland

Abstract

Many social media sites such as Del.icio.us, Flickr, or weblogs have recently become popular. This has led to adoption of tagging functions on traditional web sites at a steady pace. However, producing tagging data from these sites without supporting the social exchange involved can be regarded as an incomplete set of metadata. Although tagging captures our individual conceptual associations, the tagging system itself does not promote a social transmission that units both creators and consumers. To achieve social transmission environments for tagging, we need a formal conceptual model to represent the tagging activity and a service platform to encourage its exchange and interoperation.

1   Introduction

The label *tagging* has been applied to a fast-growing number of web sites where content has a tag that is primarily created by users themselves. Tagging is used in many social media applications such as weblogs, social bookmarking, and social networking applications. While the primary purpose of tagging is to help users organize and manage their resources, collective tagging data is used to organize and retrieve information via folksonomies, which are types of distributed classification (Gruber, 2008). A tag, or a labeled keyword, is a type of metadata for items such as resource links, web pages, pictures, blog posts, etc. and is primarily created not by machine agents, but by human users. A tagger, who the entity is creating tags, does not necessarily have to be an expert but may simply be the creator of an annotation or the consumer of an item. It is important to remember that resources can be tagged with as many or as few words as desired; there is no

restriction to placing objects in one category (Shirky, 2005). Tagging is a way of representing concepts, with a free-form list of keywords, by using the cognitive association techniques of a tagger without enforcing categorization (Kim et al., 2008). Both creators and consumers of tagged items can share their collection of tagging data. Since a large number of users participate in creating, adding, and sharing metadata in the form of keywords, this is regarded as a highly social and democratic process.

The term *folksonomy* means the practice and method of collaboratively creating and managing tags for the purpose of annotating and categorizing content (Mathes, 2004). The term, first coined by Tomas Vander Wal[19], is a fusion of the two words '*folk*' and '*taxonomy*' and it became popular on the Web around 2004 with social software applications such as social bookmarking or photograph annotation. For instance, some well-known implementations of folksonomies are del.icio.us (a social bookmarking system) and Flickr (a photo-sharing web site). CiteULike, using a similar approach to del.icio.us, focuses on academic articles, and there are a number of multimedia sites that support tagging such as Last.fm for music and YouTube for video. The power of folksonomies is obtained through an aggregate summary of the information that we are interested in, and this improves social reinforcement by enabling social connections and by providing social search mechanisms. Quintarelli (2005) points out that "*without a social distributed environment that suggests aggregation, tags are just flat keywords*."

2   What are the Problems of Current Social Tagging and Systems?

Although social tagging and folksonomies have a lot of advantages (visualization, navigation, etc.) to offer the different users who tag content items in social media sites, critical drawbacks with current tagging systems are that 1) there is no formal conceptualization in order to represent tagging data in a consistent way (Kim et al., 2008), and 2) there is no interoperability support for exchanging tagging data among different applications or people (Gruber, 2008). The simplicity and ease-of-use of tagging leads to a lack of precision with keyword ambiguity caused by misspelling certain words, singular vs. plural, synonyms, morphologies, or too-personalized tags (Golder and Herbermann, 2006; Halpin et al., 2006; Marlow, 2006). Since there are many different manners of using tags, one may not be able to understand what a given tag is about. These limitations come from a lack of standards for tag structures and little semantics for specifying the exact meaning.Aside from these problems, social tagging systems do not provide a uniform way to share and reuse tagging data amongst users or communities (Kim et al., 2007). There is no consistent method for reusing one's personal set of tags among

---

19 http://www.vanderwal.net/random/entrysel.php?blog=1750

people or communities. Although some folksonomy systems support export functionality using their open APIs and share their data via 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 is not easy to meaningfully search, compare, or merge "*similar collective tagging data*" from different applications (TagCommons, 2007).

With the usage of tagging systems increasing daily, these limitations will become critical. To overcome the limitations of current tagging systems, we need to look at an open platform for tagging similar to OpenSocial[20] that provides a common set of APIs for social networking applications across multiple web sites.
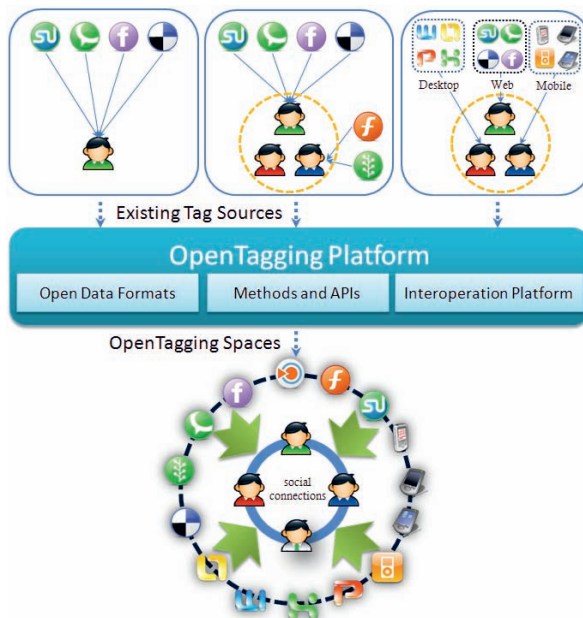

Figure 1. The OpenTagging Platform

We can see three different scenarios of using tagging data from existing tag sources in Figure 1.

- Individual perspective: Users participate in diverse social media sites by contributing to tagging activities. Although they are able to collect the tagging data resulted from these activities, the real challenge is to integrate and combine this data into a comprehensive personal view.

---

[20] http://code.google.com/apis/opensocial

- Community perspective: On the side, users are part of different communities and projects, and interact with the members of these communities by sharing or exchanging tagging data between them. In this setting, a new issue arises, i.e. the reuse of the data across multiple communities.
- Heterogeneous environments: Regardless of the individual or community perspectives, tagging data can be produced in several environments, like on the Desktop, on the Web, or even on the Mobile Web. Even so, we want be able to reuse our set of tags indifferently of the environment. However, there is no consistent method for interoperation of tagging data amongst different environments.

3   Components of the OpenTagging Platform

The goal of OpenTagging aims to make tagging data open, more universal, and apply it across any number of social tagging sites. Through continuously user participations on the platform, users can make their customized folksonomies to organize their data by their needs and interests. The interaction of diverse objects such as users, tags, and resources on the platform brings emergent semantics of tagging data and leverages social connections among participants. In order to allow users and developers to implement the social capabilities underlying tagging data, the platform consists of the open data models, the export and sharing methods, a consistent platform for interoperating one's personal set of tags between either web-based systems, desktop, and mobile applications, or for transferring tags among the desktop, the web, or the mobile.

3.1 Open data formats

These aim for specifying tagging data in a formal way. The data formats for common conceptualization of tagging data can be represented by an ontology to make a minimal commitment. A conceptualization of tagging data and activities is called 'Tag ontology' and there are some implementations using OWL (Kim et al., 2008). It is also important to note that some classes and properties from well-known RDF vocabularies (SIOC, FOAF, and SKOS etc) can be used to represent tagging activities. This approach can be considered as a method to enhance semantically links of tagging data.

3.2 Methods

In order to collect, share, or exchange tagging data, or create a bridge among heterogeneous social tagging sites, methods should implement by types of mashups. In general, most social media sites offer open APIs to expose their data and we can gather the data using them. It, however, is hard to integrate and interoperate data through diverse applications on syntactic and structural means; we need

semantic techniques such as SA-REST (Semantically Interoperable and Easier-to-Use Services and Mashups) (Sheth et al., 2007).

3.3 Platform

The Semantic Web is a useful platform for linking, exchanging, and interoperating supports on tagging data collected from heterogeneous social tagging sites (Breslin & Decker, 2007). The platform supports a social ecosystem that interlinks among objects such as individual and individual, individual and communities, or individual and the tags themselves and leverages social connections based on tags. In addition, the platform allows users to reuse and exchange tag data between people across different sources (systems) in existing social networks, which could be used to connect people who may have a common interest, or set of interests.

4   Open Data Format for Describing Tags: Social Semantic Tags

The SCOT ontology aims to describe the structure and the semantics of tagging data and to offer social interoperability of the data among different sources. Tagging is an activity or a process in which a tagger 'assigns' some tags he or she 'creates' or 'uses' on some resources. In order to represent this activity, the model represents tags clouds, the tags themselves, the resources that are being tagged, and the users that create these tags. The model also describes the properties of the tags, including their occurrence frequencies, and other tags that are used in conjunction with them. 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 enables 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. Figure 2 gives a detailed example of a tagging activity describing by SCOT instance. The Tagcloud class consists of metadata related to tagging activity such as taggers, sites, and creators and of statistical information to describe overall tag usage such as total posts, total tags, or total frequency of tags in a site. The Tag class describes a concept of an individual tag. This class includes many properties to represent the semantics (scot:acronym, scot:synonym, scot:spellingVariants, etc) and numerical features (scot:ownAFrequency and scot:ownRFrequency, etc) of a tag. The Cooccurrence class describes co-occurring tags and the co-occurring frequencies among tags.

SCOT aims to incorporate and reuse existing vocabularies as much as possible in order to avoid redundancies and to enable the use of richer metadata descriptions for specific domains.
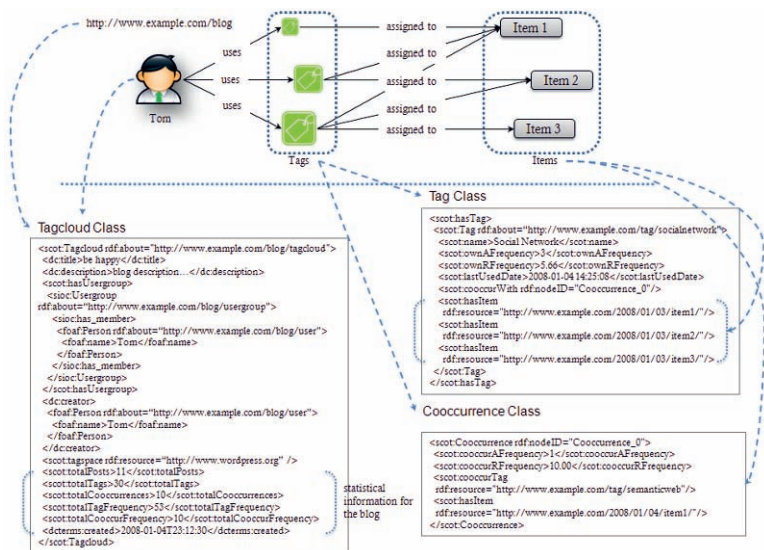
194

Figure 2. An example of SCOT instance. The SCOT models tagging activity for typical online communities including taggers, tags, items, and these relationships.

This ontology model has been made with a number of vocabularies including DC (Dublin Core Metadata)[21], FOAF (Friend-of-a-Friend)[22], SIOC (Semantically Interlinked Online Communities)[23], and SKOS (Simple Knowledge Organisation Systems)[24]. Figure 3 illustrates the relationships among these vocabularies.

4.1 How can users create SCOT instance?

We do not force any burden on users in relation to creating the semantic data and do not expect users to understand 'what the Semantic Web is' or 'what an ontology is'. We have provided SCOT Exporter[25] which automatically create semantic metadata from a set of tagging data. For instance, the SCOT Exporter for WordPress, which is a plug-in, allows the production of SCOT instance data from a certain blog. This Exporter is activated in the plug-in menu on the WordPress administration panel and it requires no user configuration in order to work. The instance created by the Exporter is located in '*http://yourhost/scot/scot.rdf*', when

---

21 http:// dublincore.org

22 http://foaf-project.org

23 http://sioc-project.org

24 http://www.w3.org/2004/02/skos

25 http://scot-project.org/applications/wp-exporter/

tags are changed in the blog, the instance is dynamically updated. The initial version for the export has developed based on the assumption that categories in WordPress are used as tags. We also offer the exporter for the Ultimate Tag Warrior[26], a popular and powerful WordPress plugin, which allows a user to add tags either through the Write Post page in WordPress in a tag box.

4.2 How can we provide interoperation amongst different sources?

To realize the OpenTagging platform, we make it possible to exchange, compare, and integrate tagging data across different applications or sources and to offer interoperation in the tag spaces. Although a user can create a SCOT instance data set using a SCOT Exporter from a single online community such as a weblog, the Exporter provides a simple method for exposing a SCOT instance without interoperation mechanisms. Thus we need a method for sharing and interoperating this semantic metadata.

5   The int.ere.st Web site and its Methods

int.ere.st[27] is a web site 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 (Kim et al., 2007). The site (see Figure 4) 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, social web sites can be stored in a repository as SCOT instances 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 and these tagging data sets 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. The following are some of the main methods implemented in the site.

The screenshot (fig. 2) shows the search results for the tag 'ontology.' In this example, the left-hand side shows SCOT instances with associated detailed information such as top tags, the creator, number of members and items, total tags, total

---

26 http://www.neato.co.nz/ultimate-tag-warrior/

27 http://int.ere.st

co-occurrence tags, and tag spaces. If a user clicks on a title in the search results, the right-hand side visualizes a tag cloud for the selected SCOT instance with related items.
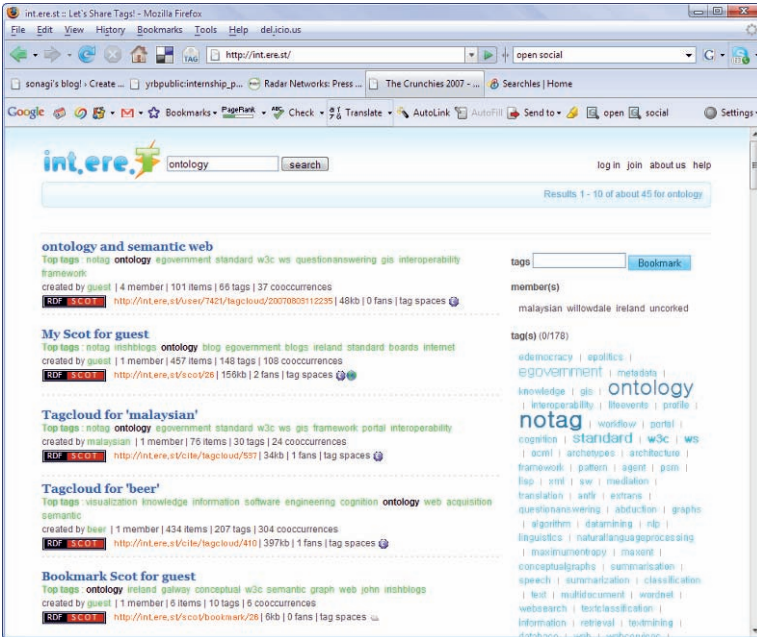

Figure 3. int.ere.st web site

5.1 Aggregate

A user can gather a collection of tagging data that he or she has assigned to resources in distributed applications. For instance, the site can aggregate bookmarks, images, and videos with tags from del.icio.us, Flickr, YouTube, or other online applications using their open APIs. The collected data is automatically transformed into semantically structured data that includes the relationships among users, tags, and resources. If tagging data is already created by the SCOT exporter in a certain blog, a user can directly import the instances from their site. We also provide an importing method in which users can import their SCOT instances from a file or URL. Then, the aggregator for the collected or imported instances runs periodically and automatically. This is a first step for sharing tags from different resources in different tag spaces.

5.2 Search & Browse

There are several ways to search tag information on the site. Firstly, a tag search allows users to look for similar patterns of tagging or persons with related interests based on tags. Secondly, a user can search for tags or resources using SPARQL-

based semantic search methods with these search operators: "and", "or", co-occurring tags, and broader or narrower relationships. These operators enable users to restrict their search conditions. Thirdly, when the '*created by*' field from the search results is clicked for a specific SCOT instance, all SCOT instances created by the creator are listed. This will help users find interesting new people in the system, much as a user refers to instances to find interesting new ones.

5.3 Bookmark

We also provide a bookmarking and tagging method for each SCOT instance so that a user can participate in the tagging activity and share experiences with other people. If a user is interested in tagging data from a certain instance, he or she can create a bookmark, with tags, for the instance. We provide '*fans*' as a concept for a list of such people; when someone has added a certain SCOT as a bookmark, a fan connection is created. Social connections can be made with other individuals interested in just about any topic. In addition, a user can take advantage of all the work other people have done. A list of bookmarked instances is located in the "my interest" menu.

5.4 Share

The site exposes various and structured types of user contributions in the system and also connects to other sources of data using Semantic Web technologies. For instance, personal information can be exposed as FOAF and SCOT instances in the system can be mapped into SIOC. The SCOT ontology can be classified with several types such as "imported", "bookmarked", and "integrated" one in the system. The bookmarked type is created by other users; the integrated type (that is a merging of at least two instances) is created by the logged in user; the imported ontology can be either of the two. The bookmarked type is described using the property from FOAF and the integrated type is mapped to the FOAF maker property. In addition, all types of SCOT instances for a certain user are mapped to the Item class from SIOC. This process can be done automatically. The mapping among SIOC, FOAF and SCOT together provides a way to enhance social connections that are distributed and shared among people.

6   int.ere.st as a Platform

int.ere.st is the first OpenTagging platform for 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 via the underlying SCOT ontology. Those functionalities help users exchange and share their tagging data based on 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. The RDF vocabularies can be interlinked with the URIs of SCOT instances that are generated in the site and shared in online communities.

7   Conclusions

We discuss the OpenTagging platform for interoperation of social tagging data. The platform allows users to reuse and exchange tagging data between people across different sources (systems) in existing social network, which could be used to connect people who may have a common interest, or set of interests. Although it's still in an early stage, we hope additional effort will make the OpenTagging platform more practical and useful. We expect that the SCOT project (http://scot-project.org) provides open discussions for community and the int.ere.st as a test-bed continues to bring novel approaches and solutions to problems in social tagging and interoperation processing.

References

Breslin, J.G. and Decker, S., (2007). The Future of Social Networks on the Internet: The Need for Semantics, IEEE Internet Computing, vol. 11, no. 6, pp. 86-90, Nov/Dec.

Golder, S. and Huberman, B. A. (2006). The Structure of Collaborative Tagging Systems. Journal of Information Sciences. 32(2). 198--208.

Gruber, T. (2007). Ontology of Folksonomy: A Mash-up of Apples and Oranges. Intl Journal on Semantic Web and Information Systems. 3(2).

Gruber, T. (2008). Collective knowledge systems: Where the Social Web meets the Semantic Web. Journal of Web Semantics 6(1). 4-13.

Halpin, H., Robu, V. and Shepard, H. (2006). The Dynamics and Semantics of Collaborative Tagging. In Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW06).

Kim, H. L., Yang, S. K., Breslin, J. G. and Kim, H. G. (2007). Simple Algorithms for Representing Tag Frequencies in the SCOT Exporter. in IAT. IEEE Computer Society. pp. 536-539.

Kim, H.L., Passant, A., Breslin, J., Scerri, S., Decker, S. (2008). Review and Alignmnet of Tag Ontologies for Semantically-Linked Data in Collaborative Tagging Spaces, In Proceedings of the 2nd International Conference on Semantic Computing, San Francisco, USA.

Marlow, C., Naaman, M., Boyd, D. and Davis, M. (2006). HT06. tagging paper. taxonomy. Flickr. academic article. to read. in HYPERTEXT 06: Proceedings of the seventeenth conference on Hypertext and hypermedia. ACM Press. New York. NY. USA. pp. 31--40.

Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Retrieved June 25. 2008. from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Quintarelli, E. (2005). Folksonomies: power to the people. Retrieved June 25. 2008. from http://www.iskoi.org/doc/folksonomies.htm.

Sheth, P.A., Gomadam, J., Lathem, J. (2007). SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups, IEEE Internet Computing, vol. 11, no. 6, pp. 91-94, Nov/Dec.

Shirky, C. (2005). Ontology is Overrated: Categories. Links. and Tags. Retrieved 25 June. 2008. from http://www.shirky.com/writings/ontology-overrated.html.

TagCommons, (2007). Functional Requirements for Sharing Tag Data, Retrieved 25 June, 2008, from http://tagcommons.org/2007/02/28/functional-requirements-for-sharing-tag-data/.