

Association for Information Systems AIS Electronic Library (AISeL)

AMCIS 2010 Proceedings

Americas Conference on Information Systems
(AMCIS)

8-2010

Why Do Folksonomies Need Semantic Web Technologies?

Hak-Lae Kim

Samsung Electronics Co., LTD. 416, Maetan-3dong, Yeongtong-gu Suwon-city, Gyeonggi-do, 443-742 Korea,
haklae.kim@gmail.com

Jae-Hwa Choi

Dankook University, jchoi@dankook.ac.kr

Follow this and additional works at: <http://aisel.aisnet.org/amcis2010>

Recommended Citation

Kim, Hak-Lae and Choi, Jae-Hwa, "Why Do Folksonomies Need Semantic Web Technologies?" (2010). *AMCIS 2010 Proceedings*. 68.
<http://aisel.aisnet.org/amcis2010/68>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Why Do Folksonomies Need Semantic Web Technologies?

Hak-Lae Kim

Samsung Electronics Co., LTD.
416, Maetan-3dong, Yeongtong-gu
Suwon-city, Gyeonggi-do, 443-742
Korea
haklae.kim@gmail.com

Jae-Hwa Choi

College of Business
Dankook University
San#29, Anseo-dong, Dongnam-gu, Cheonan-si,
Chungnam, 330-714, Korea
jchoi@dankook.ac.kr

ABSTRACT

This paper is to investigate some general features of social tagging and folksonomies along with their advantages and disadvantages, and to present an overview of a tag ontology that can be used to represent tagging data at a semantic level using Semantic Web technologies. Several tag ontologies have been developed with a specific purpose and used in various websites. However, in order to represent tagging data at semantic level existing tag ontologies need to be interlinked, since individual tag ontology cannot represent overall features of tagging activities. After introducing conceptual overview of tagging and folksonomies and tag ontologies, we will propose the combinational model for linking tag ontologies.

Keywords

Folksonomies, Semantic Web, Ontology, Tag Ontology.

INTRODUCTION

The idea of tagging is simple: people are tagging the content they create in public spaces with free-text keywords, and the collection of many people's keywords can serve as useful metadata when aggregated (Gruber, 2007). Based on this principle, a growing number of tagging applications have begun to provide users with the ability to connect to other users via their own keywords. Tagging plays an important role in the emergence and production of an incredible amount of distributed information by communicating, collaborating, and sharing tagged resources amongst users. With this new phenomena, classification on the Web is now changing in that online resources can easily be classified in many different categories beyond the types we encounter in the physical world (Weinberger, 2005). The results of tagging are emerging as collective semantics created through active user participation. The increasing interest in tagging leads to new challenges in order to improve the findability of resources and connectivity between people. Although tagging is an evolving and "bottom-up" approach based on users' contributions, representing the social connectivity between the entities that make up tagging activities has to be discovered and improved. For instance, detractors of tagging refer to a lack of precision and low findability as major limitations which prevent the sharing of tagging data (Mathes, 2004). We believe that we can leverage this new model of tagging content on the Social Web by combining it with alternative but complementary technologies such as the Semantic Web.

The objectives of this research are to 1) explore overall features of tagging and folksonomies, including definitions, advantages and disadvantages, 2) propose a combinational model based on tag ontologies to make tagging data unambiguous and to be shared tagging activities across heterogeneous sources. The contributions can be summarised: 1) identify the differences between folksonomies and ontologies, 2) address the reasons that folksonomies can be represented by ontologies, 3) propose a combinational model by interlinking existing tag ontologies in order to represent tagging data at a semantic model. In the following sections, we will begin by clarifying the various concepts and terminologies related to tagging, and then we will discuss the distinct roles that are present in tagging processes.

SOCIAL TAGGING CONCEPTS

Definition

A tag is a metadata keyword or a labeled term associated with a resource such as a link, web page, picture, or blog post, etc. Previously, research literature on tagging dealt primarily with online resources such as blog posts (Mathes, 2004;

Merholz, 2004; Mejias, 2005; Kroski, 2005; Vander Wal, 2005). Tags are in general chosen by people (e.g. the content creator or author, a consumer or reader, etc.), and are based on their understanding of a resource according to their own perceived utility, where they are not chosen from existing controlled vocabularies or strict taxonomies (Mathes, 2004). The idea of tags is that they reflect the end user's thoughts or intentions by using informal personally preferred terms. While users who create tags do not necessarily have to be experts, the terms as keywords, categories, or classifications for the resource hold an intended meaning for the users (Golder and Huberman, 2006). Once tags are assigned to a resource, different users or applications can use and share them in several ways. In this sense, tags can be seen as a simple way to allow users to interact with a site or a service based on a richer experience and a sense of participation (Mathes, 2004; Kroski, 2005).

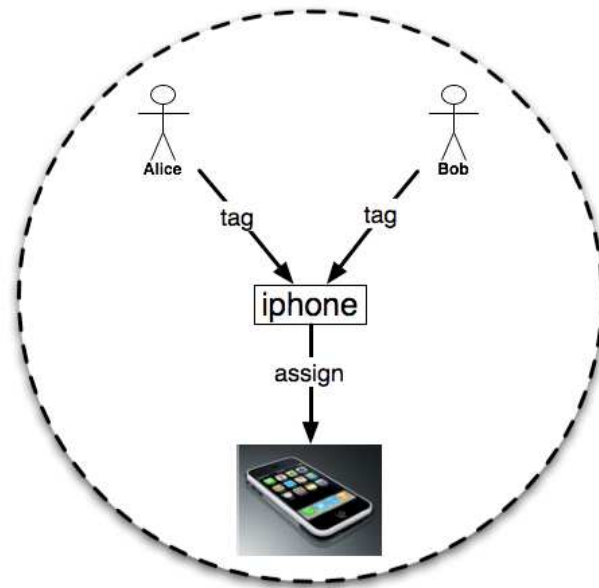


Figure 1 Tagging model: People tag a specific resource (e.g. a photo) with term(s).

Tagging is the overall activity of contributing to a collection of tags by users. As shown Figure 1, tagging requires people to participate in tagging activities for their own specific reasons and thus the entity of this process includes users, tags, and resources as well as their relationships. In this perspective, the tagging process goes with an implicit agreement of sharing tagging entities amongst users, rather than being just a process of sharing resources. A folksonomy is a type of distributed classification system and is formed from a collection of keywords created by users. This can be defined as a social, shared categorisation system influenced and consumed by or across users, reflecting varied perspectives in a collaborative and decentralised way (Golder and Huberman, 2006; Marlow et al., 2006; Halpin and Shepherd, 2007). Thomas Vander Wal, the person who coined the term, emphasises “user-generated classification, emerging through bottom-up consensus” (Vander Wal, 2005). He asserts that a folksonomy, as a fusion of the words ‘folk’ and ‘taxonomy’, is composed of three important components - the tags themselves, the resources that are being tagged, and the users that create these tags. A real large-scale application of folksonomy usage appeared with Delicious in late 2003, and this was subsequently adopted by various other social software sites.

According to Vander Wal (2005), folksonomies are distinguished by two typologies: broad and narrow folksonomies. The former refers to where many people are describing the same object with their tags in a system such as Delicious. In a tagging space, everyone can tag an object according to their interests, even if the methods of tagging used among them is different. Thus, the power law of tagging behaviours observes that most tags appear only once or rarely in a small groups or few users, whereas a few popular tags are often used by many users (Quintarelli, 2005). Broad folksonomies, following a variety of levels of user participation, are a way of investigating the benefit of the network effect and other trends in tagging. In contrast to broad ones, narrow folksonomies mean that a small number of users will tag individual objects (e.g. as is the case with Flickr).

At present, tagging on a large number of social software sites has become quite popular, and tagging in traditional web sites (e.g. message boards) is being adopted at a steady pace. Some well-known implementations of folksonomies are Delicious and Flickr. The former is a social bookmarking system where users tag and store web links, which they want to find again in the future, while the latter is a photo-sharing web site where users upload and tag their own photos to aid

retrieval by themselves and others. CiteULike¹, using a similar approach to Delicious, focuses on academic articles, and there are number of multimedia sites that support tagging such as Last.fm² for music and YouTube³ for videos. From an information visualisation perspective, tag clouds are the most popular visual interface to represent multiple tags and for navigating resources. Typically tags displayed in a tag cloud on most sites are differentiated by size, font or color depending on how many times the tag has been used or how many resources have been assigned with that tag (Hayman, 2007).

Social Tagging and Its Terminologies

In any emerging field, a new phenomenon will have a corresponding set of new terms that describe it or are related to it, and there are numerous definitions for the tagging phenomenon as a method of categorising information. The term folksonomy is emerging as the dominant term: there is however a lack of clarity when talking about tagging and folksonomies, as these terms are often used interchangeably to describe the same phenomenon. As well as folksonomy, some of the other terms that have been proposed: *ethnolocation* (Merholz, 2004), *distributed classification* (Mejias, 2005), *social classification* (Smith, 2004), and *communal categorisation* (Shirky, 2005). Speller (2007) focuses on the social and collective features of the tagging process itself as opposed to the taxonomic aspects.

The relationship between tagging and a folksonomy is not clear if a distinction between personal and social tagging is made. Whereas tagging as a process is part of folksonomy creation, a collection of tagging instances on its own is not a folksonomy. The result of aggregating tags or items from individual tagging instances will not be a folksonomy either. That is, tagging is not only a way of representing concepts by cognitive association of individual users, but also a social and democratic process to encourage social connections amongst users. Thus the benefits of tagging can be differentiated on both the individual and the social perspective. The former point of view represents a very powerful personal organisation and retrieval, because the tagging process in this case is primarily dependent on a user's intentions. Users can assign many layers of meaning to resources and can navigate the resources when they want to find one. Social perspective plays a significant role to encourage implicit collaboration amongst users or to reflect a collective memory that represents shared knowledge. Thus, the latter can be called social tagging.

Despite terminology debates, various terms tend to assume that a continuous social transformation, from individual to collaborative tagging based on user participation, leads to unexpected emergent semantics and collective intelligence. As Quintarelli (Quintarelli, 2005) points out, without a social distributed environment that allows for aggregation, tags are just flat keywords. By combining a collection of tags with this transformation, folksonomies can emerge. Compared to the tagging process, folksonomies tend to be dependent on a particular system. Based on the system, folksonomies have different characteristics defining tagging entities and their relationships, and also may show different views and contexts of their resources. Folksonomies can be defined by the following features:

- An aggregation of tags for multiple items between multiple users
- The users must be able to see and be influenced by other's tags in the tagging process (i.e. creation and reuse of tags).
- A combination of methods to support adding, managing, and sharing etc.

For the remainder of this paper, we will use the term *Folksonomy* to refer to a certain system enabling distributed classification, whereas we will use the term *Tagging* to identify the activity of assigning a set of tags to a resource. A *Tag Cloud* is an interface of displaying collection of tags or folksonomies, and not a folksonomy in itself. For instance, Delicious allows users to organise their bookmarks at a personal level (i.e. tagging), and at the same time the aggregated tag behaviours (i.e. folksonomy) is represented by the tag cloud as shown in Figure 2.

¹ <http://citeulike.org>

² <http://lastfm.com>

³ <http://youtube.com>

Tag Cloud: Popular

KEY: blue tags are tags you have in common with everyone else.

Sort: Alphabetically | By size

.net 2008 3d advertising ajax and animation api apple architecture art article articles artist audio **blog** blogging **blogs** book books browser **business** car cms **code** collaboration comics **community** computer converter cooking cool **css** culture **data** database **design** desktop **development** diy documentation download downloads drupal ebooks economics **education** electronics email entertainment environment fashion fic film finance **firefox** **flash** flex flickr food forum free freeware fun funny **gallery** game **games** geek **google** government graphics green **guide** hardware health **history** home hosting house **howto** html humor icons illustration images imported information inspiration interactive interesting internet **iphone** japan **java** **javascript** jobs jquery kids language learning library linux list lists literature mac magazine management maps marketing math media microsoft mobile money movie movies mp3 **music** network **networking** news online opensource osx people phone photo photography photos photoshop php plugin podcast politics portfolio privacy productivity **programming** psychology python radio rails realestate recipe recipes **reference** religion research resources reviews **rss** ruby rubyonrails school **science** search security seo shop **shopping** social socialnetworking **software** statistics streaming teaching tech technology tips todo **tool** **tools** toread travel **tutorial** tutorials tv twitter typography ubuntu usability **video** videos vim visualization web web2.0 webdesign webdev wiki wikipedia windows wishlist wordpress work **writing** youtube

Figure 2 Example of a tag cloud in Delicious

Characteristics of Tagging

The essence of tagging is the “voice of people” reflecting the “long tail” interests of individual users to categorise a resource without cultural, social, or political bias (Kroski, 2005). Because they do not need to understand predetermined terminologies and hierarchical structures in traditional classification schemes, the overall process of adding and consuming tags is quick, simple, and straightforward. Moreover, tagging has a very small learning curve, since users can express the direct information needs and desires they have with their vocabularies (Kroski, 2005). Tags are in accord with the user’s own words, not imposed by a system (Neal, 2007), i.e. they reflect the user’s mental concepts. In tagging systems any type of resource can fit into multiple categories, and these give us the chance to observe the resources via multi-faceted or multi-dimensional features. For example, a photo on Flickr could have a collection of tags attached such as birthday, family, or holiday. In this way tags are useful entry points to explore resources through faceted browsing by searching, filtering, navigating, and exploring other user’s tags and tagged items.

Tagging is usually, done by people consuming resources in a social environment (Vander Wal, 2005). Although an individual’s tags tend to be very subjective, individuals by nature have a spirit of sharing their tags on communities or public spaces. Because everyone has a common goal of categorising their own information and also sharing it with others, tags become social. In this way folksonomies, as the collective result of collaborative tagging, give us an opportunity to make social connections based on tagging behaviours and patterns. Many tagging applications allow users to create networks with other users that they have identified as interesting. This feature fosters the development of communities around similar interests and viewpoints. Despite its low cognitive cost, its capability to match the user’s real needs and language, and its value in serendipitous browsing, tagging by nature has a number of important issues regarding tag quality. Amongst others we mention intrinsic imprecision and lack of visualisation tools to enable users to navigate through the mass of tags (Quintarelliet al., 2006). The simplicity and ease-of-use of tagging leads to problems concerning effective categorisation and information retrieval. The major drawbacks of tags may be the lack of precision. Tags are not normalised for synonymy, morphology, and there can be many different manners of specifying the exact meaning such as capitalisation, singular vs. plural, or the relationships between tags. Consequently, free indexing is often considered to be of poor quality. The problem with visualisation is that flat tag clouds are not sufficient to provide a semantic, rich and multidimensional browsing experience over large tagging spaces (Quintarelliet al., 2006). In such a structure, the most popular tags are usually displayed through an alphabetically ordered list in which the font size increase with the tag’s relevance.

COMPARISONS BETWEEN FOLKSONOMIES AND FORMAL TAXONOMY

Categorisation and classification

The terms categorisation and classification are often confused and used interchangeably. There are obvious similarities between them in that they are mechanisms for organising information or things into categories (Jacob, 2004; Mathes, 2004). However, strictly speaking there is a subtle difference between them, and the distinction leads to meaningful differences for using them within information environments (Jacob, 2004) and the design of these environments.

A categorisation is a “synthesis of similarity” in the process of creating categories (Mathes, 2004), whereas a classification is a systemic arrangement in groups or categories according to pre-defined criteria (Jacob, 2004). Categorisation is the process of grouping entities by similarity - one of the most central and generic ideas of conceptualisation (Bailey, 1994). Any kind of organisational scheme is a categorisation or classification: for example, sorting cars by size, documents by type, or languages by country. Things are grouped together in both systems according to their similarities. However, a categorisation does not impose an order between categories, whereas a classification defines an order between categories or classes and describes their relations within a hierarchical structure. Thus membership in categorisation is non-binding. By contrast, classes in classification are mutually exclusive and non-overlapping. In tagging systems the set of tags is flat, without any ordering and with an overlap between multiple resources.

Formal classification: taxonomy and ontology

Taxonomy is the organisation of a particular set of information for a particular purpose in a hierarchical structure (Gilchrist, 2003; Garshol, 2004). Taxonomies define exact names, known as *taxa* (singular taxon), for everything and show which things are parts of other things: often called parent-child or broader-narrower relationships. One of the most widely known taxonomies is the Dewey Decimal Classification (DDC), which is a method of categorizing books in a library by subjects. The DDC was created by Melvil Dewey in 1876 and has since been extensively used in libraries (with modifications and extensions). Taxonomies are often created to describe categories and subcategories of topics on the Web. For instance, the Yahoo! Directory⁴ or the DMOZ open directory⁵ are typical examples providing directories or categories of web resources in a fixed hierarchical structure. An Ontology is a set of well-defined concepts describing a specific domain. The term as a philosophical sense refers to a “systemic account of existence” (Gruber, 1995) and is “the study of the categories of things” within a certain domain (Sowa, 1999). The term ontology in computer science was borrowed from artificial intelligence, and ontologies have been closely linked with knowledge representation and logical inference techniques (Garshol, 2004). Within this area, an ontology refers to “formal, explicit specifications of shared conceptualisations” at a semantic level (Gruber, 1993). According to Sowa (Sowa, 1999), ontologies can be classified as terminological (e.g. WordNet) ontologies that are not fully specified by axioms, and axiomatised ontologies (e.g. Cyc Upper ontology) (Fishcher, 1998).

Table 1 summarises the characteristics of categorisation, taxonomies, and ontologies. Taxonomies and ontologies have similarities to eliminate ambiguity of terms and represent hierarchical structures. The distinction between ontologies and taxonomies is sometimes vague. A very simple ontology without properties and constraints could be called a taxonomy. However, ontologies have strict and formal rules for describing relationships among concepts and for defining properties. That is, an ontology is a set of “formal explicit descriptions in a domain of discourse (i.e. classes)” and it specifies relation types (taxonomic and hierarchical relations of classes) organised in a partial ordering (Fishcher, 1998 and Noy and McGuinness, 2001). While in taxonomies simple properties can be attached, defining a chosen hierarchy, in general ontologies have stronger expressivity than taxonomies. In addition to the expressivity of ontologies, different benefits can be utilised from ontologies. Noy and McGuinness (2001) stated 5 reasons to use ontologies: 1) to share common understanding of the structure of information among people or software agents, 2) to enable reuse of domain knowledge, 3) to make domain assumptions explicit, 4) to separate domain knowledge from the operational knowledge, 5) to analyse domain knowledge. According to Gruber (Gruber, 1993), ontologies are not limited to roles of taxonomic classification (ontology-as-specified conceptualisation) and centrally controlled categorisation (a very specific methodology for agreeing on a conceptualisation). He emphasises that ontologies are a way of knowledge exchange and reuse, differentiating them with taxonomies. For example, ontology languages such as RDF and OWL provide a more precise description of domain knowledge and they also facilitate machine-processable knowledge sharing and exchange.

⁴ <http://dir.yahoo.com>

⁵ <http://www.dmoz.org>

	Categorisation	Taxonomy	Ontology
Organisation of information	X	X	X
Elimination of ambiguity		X	X
Synonym control		X	X
Hierarchical structure		X	X
Properties and axioms			X
Sharing	X	X	X

Table 1 Features of categorisation, taxonomy, and ontology

Contrasting folksonomy and formal classificatons

There are a number of debates currently on the advantages and disadvantages of both folksonomies and traditional classifications such as taxonomies or ontologies (Mathes, 2004; Weinberger2005; Shirky, 2005). As shown in Table 2, the folksonomy and classification systems can be placed at the two opposite ends of the classification spectrum. Folksonomies are an emergent set of tags with a completely flat, non-hierarchical system. No parent-child relationships are directly specified between these tags (Mathes, 2004; Weinberger, 2005; Golder and Huberman, 2006). Folksonomies are generally unsystematic and are not constrained by a requirement for mutual exclusivity. Thus, assignment of tags is multi-faceted and massively multidimensional in a tagged item (Vander Wal, 2005; Quintarelli, 2005). In contrast, formal classifications are generally a hierarchical structure of well defined, mutually exclusive, and non-overlapping classes defined in a super-sub relationship (Jacob, 2004).

	Folksonomy	Formal classification
Background	Web communities	Metaphysics, AI, knowledge engineering
Structure	Flat structure with clusters of entities	Hierarchical structure of fixed vocabularies
Formality	None	High
Creation	Community users	Domain experts
Ambiguity control	None	Control possible
Creation cost	Low, created by users	High, created by experts
Change	Highly dynamic, changes constantly	Rigid, often has to be recreated to accommodate change
Usability	No expertise required	Requires proficiency in handling
Vocabulary	Personal vocabulary	Pre-defined or controlled vocabulary
Scalability	Expansive	Restrictive

Table 2 Comparison between folksonomy and taxonomy

The differences in terms of controlled vocabularies can be used to differentiate between them. Folksonomies offer uncontrolled and personal vocabularies. It means that everyone can choose tags to appropriately describe items from an individual standpoint, because of no global consistency to use a particular term. Folksonomies encourage users to describe their mental concepts in a decentralised and collaborative way (Quintarelli, 2005), but there can be many different words for the same thing. Thus, the meaning of or between tags has to be interpreted by analysing the relationships between users, tags, and resources in a social context. By contrast, classifications are in general comprised of controlled or pre-defined

vocabularies. The classification systems manage the use of synonyms, homonyms, lexical anomalies and so on. This is the advantage for retrieval, navigation and interoperation of information on the Web. However, the main disadvantages of controlled vocabularies are that they quickly become outdated and non-experts or less trained professionals are restricted to use the classifications. As Shirky (Shirky, 2005) stated, few experts studying a particular domain can contribute to make traditional classifications using a hierarchical taxonomy. Furthermore, it is important to note issues regarding the development and the management of both systems. The cost of creating traditional classification schemes is high and it also becomes expensive to manage and update them, whereas tagging systems provide a low alternative for categorising information (Mathes, 2004; Quintarelli, 2005). However the overall cost of managing folksonomies might not be low because the need to overcome the drawbacks discussed above comes at an additional costs. It is true that folksonomies provide the capability to match the user's way of thinking and organising information, but their full-fledged power is still challenged.

LIMITED SOCIALITY OF FOLKSONOMIES

The most important issue associated with tagging is how we can extract a meaningful experience from the users' tagging practices and how we can provide social environments to encourage social connections from folksonomies. There are many challenges to achieving these goals.

As we discussed above, tags by nature have various limitations regarding semantics and structures. Even if social software sites are controlled to be able to eliminate tag ambiguities, they are limited to handle the input, output, and matching of tag phrases (Gruber, 2007). In addition to these inherent drawbacks, the exchange of tagging data is more difficult to achieve. Tagging practices that users and communities create on one site or application are often locked into that particular site, and cannot be reused in a different system or by a different user. As social software sites have rapidly increased, users who participate in tagging activities in different locations encounter significant problems when they try to re-use a collection of tags. The problems of current folksonomies are classified as follows:

- Individual perspective: they may build up a rich collection of tags on a certain site or application over time, but the size and scope of this collection will depend on how regularly the site or application is being used. When beginning to tag content on a new system, the user may have certain amount of reluctances to create new tags if they do not expect to be using the system that much, and also they may not wish to recreate their existing tag information on other sites (Kim et al., 2010).
- Community perspective: users are part of different communities and projects, and interact with the members of these communities by sharing or exchanging tagging data between them (Kim et al., 2008a). The idea that tagging data can be interoperated with other's data is widely accepted (Gruber, 2008). In this setting, a new issue arises, i.e. the reuse of the data across multiple communities. We have real challenges at different levels of interaction among the communities (Tagcommons, 2007).
- Heterogeneous tag formats: Tagging systems do not have a uniform structure and no agreed-upon semantics to share, exchange, and reuse tag metadata amongst users or communities (Kim et al., 2008a). Also, there is no consistent method for reusing one's personal set of tags between either web-based systems, desktop applications, or for transferring tags among different computing environments. Because of these reasons it is not easy to meaningfully search, compare, or merge similar collective tagging data and it is also difficult to improve social connections of people based on the tagging behaviours (Gruber, 2007; TagCommons, 2007).
- 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 to be able to reuse our set of tags independently of the environment. However, there is no consistent method for interoperation of tagging data amongst different environments.

While tagging is an evolving, bottom-up approach based on user's contributions, socialising the results of tagging have yet to improve in a manner that supports discovery, and sharing of tagging data. With the use of tagging systems increasing daily, these problems will become critical. Combining alternatives technologies such as the Semantic Web offers the possibility to address these limitations and enhance tag sharing capabilities.

BENEFITS OF SEMANTIC WEB TECHNOLOGIES

Semantic Web technologies in general allow us to expose human knowledge to machines in order to performing automatic data linking and data integration. In particular, ontologies as an enabling technology for the Semantic Web enable knowledge exchange among different users and applications by providing reusable constructs. That is, these technologies can improve

current tagging environments, not only by representing semantically tagging data, but also by allowing users to share and exchange the data across sites, services, or applications.

In this section, we will discuss why folksonomies need to adopt Semantic Web technologies and what kinds of benefits folksonomies can obtain from them. We will concentrate on the representation of tagging data using Semantic Web technologies because: 1) semantically-enhanced representations of tagging data using ontologies contributes toward provide standard forms for describing tagging entities (i.e. users, tags, and resources) including their relationships, 2) it can establish links amongst the entities including for expected and unexpected reuse of information.

Augmenting data structure with semantics

Semantic Web technologies can add structured values into tagging data and also can augment the data with semantics representing some of the entities and their relationships (Gruber, 2008). Tags have a high level of ambiguity when it comes to their meaning, due to various types of spellings and no synonym control etc. Ontological representation can allow a set of tags to define their meaning underlying their context including users, locations, and usage patterns. From a slightly different standpoint, Gruber (2007) and Spivack (2005) emphasise the need for folksonomies and ontologies to work together, aiming to identify and formalize a conceptualisation of tagging data at a semantic level. Typical social tagging systems do not provide explicit links amongst the entities, nor do they expose their data in a standard form. These approaches can be realised as a standardised metadata schema - called Tag Ontologies - representing both the structures and semantics of tagging data within some domains of interests. Furthermore, this type of ontology does not just define a meaning of a certain tag, but can also robustly represent the relationships among the entities that shape tagging activities, explicitly stating the knowledge structure of tagging data.

Enabling data sharing and exchange

As more users contribute to social software applications, they may want to be able to reuse their data across different platforms. A basic principle of social tagging is sharable features. Once a user assigns a set of tags to a resource, there is an assumption that tagging entities and their relationships are shared in a particular platform and the tags become a medium for building social relationships among users. Presently, the shared knowledge structure of tagging is locked into a site or an application and is scattered across the sites where users participated. These make a barrier to fully exploit the sociality of tagging and to provide explicit and implicit interactions on the Web.

Several challenges have to be addressed to solve these issues: tagging data needs to be described with richer semantic representations and the methods like open APIs have to be offered to share and aggregate the data. These are basic requirements for the goal realising sharable features of social tagging systems. Semantic Web technologies offer a good balance in this process with standardisations for describing and querying information (e.g. RDF and SPARQL, respectively). For instance, tags, users and their relationships in tag ontologies are expressed and in both machine and human understandable representations and at the same time accessed by both agents. These technologies provide effective ways not only to integrate tagging data across different platforms with richer semantic representations, but also to share the data among other users or communities.

Tag ontologies, supporting Linked Data principles (Berners-Lee, 2006), can publish semantic tag metadata and make it easy to interlink, discover and consume the data on the Semantic Web. To weave a web of scattered tags based on conceptual schemes describing them, tag ontologies could be viewed good practices not only for identifying tags in formal semantics but for building shared tag knowledge graphs. The Linked Data is a term used to describe a method of exposing, sharing, and connecting data on the Web (Bizer et al., 2008). Linked Data can enhance an opportunity to make resources interconnected on the Web, and be reused in unexpected ways. Berners-Lee (Berners-Lee, 2006) outlined four principles of Linked Data:

- use URIs to identify things as resources;
- use HTTP URIs for locating and dereferencing these resources;
- provide useful RDF information about the resource when its URI is dereferenced;
- include RDF statements that linked to related URIs so that they can discover related things.

Resources on Linked Data have specific links as URI to identify the resources and have representation (RDF, RDFa, or Microformats etc), which can be machine-processable. With combining the Semantic Web and Linked Data, isolated tagging data can be easily made mobile and integrated across different applications and can be accessible and re-used by both machines and human. The semantic structure of social tagging could also combine automatic inferences with the expertise of

users by integrating the validation these inferences within the natural use of the system. This approach is a good start for the sharing and exchange of separate tagging activities on different platforms.

TAGGING ON THE SEMANTIC WEB

Semantic Web technologies can provide a common conceptualisation of tagging activities not only for representing as standardised languages but also for describing the enormous diversity in social perspectives on a semantic level. Tag Ontology is to describe the relationships between an agent, an arbitrary resource, and one or multiple tags (Kim et al., 2008b). In particular, the TagCommons project proposed functional requirements for tag ontologies that aim to clarify the core concepts and related information. There are some efforts for developing tag ontologies, and core concept for these ontologies is *Tagging* that is the act of associating tags with an object or item. This concept is usually comprised of following concepts: 1) tag: a keyword or phrase that is recognizable by people and computers, 2) resource: a thing to be tagged (identifiable by a URI), 3) tagger: someone or thing doing the tagging, such as the user of an application. In other words, tag ontology is to explicitly represent these objects and the relationships between them. In this sense, the discussed limitations such tag ambiguities and sharing can be solved. So far, although there has been developed several tag ontologies, one model cannot be covered overall features of tagging activities. Thus, in this section we introduce some existing ontologies in brief and then we propose a combination of those ontologies that aims to describe collection of tagging activities across heterogeneous sources.

Figure 3 shows the combination amongst existing tag ontologies. Newman's model describes relationships between three core concepts (Tagger, Tagging, and Tag) to represent a tagging activity (Newman et al., 2005). For example, the tagger 'Alice' assigns the tag 'apple' into the resource 'Videos#id' on the date '2010-04-25'. However, since the tag 'apple' does not have a specific meaning – fruit or company, we need to describe the relationship between the tag and its meaning. In MOAT, the Meaning class which aims to allow the meaning of tags to make unambiguous, represents user-provided meanings for tags (Passant and Laublet, 2009). In addition, tags have their own occurrences that are an important factor to identify how the tags used, and these features are based on collecting whole usages of certain tags. The SCOT (Social semantic Cloud of Tags) represents collective tagging activities and provides the most appropriate representations for folksonomies (e.g. tag space, community, frequencies, tag co-occurrence, etc.) (Kim et al., 2010). For example, the TagCloud class cannot only contain tags that occurred in a source, but also can represent user groups who participated in tagging activities. The property scot:hasUsergroup links the TagCloud class to Usergroup class of SIOC. Using this approach tagging entities can be semantically represented and shared across different sites or sources.

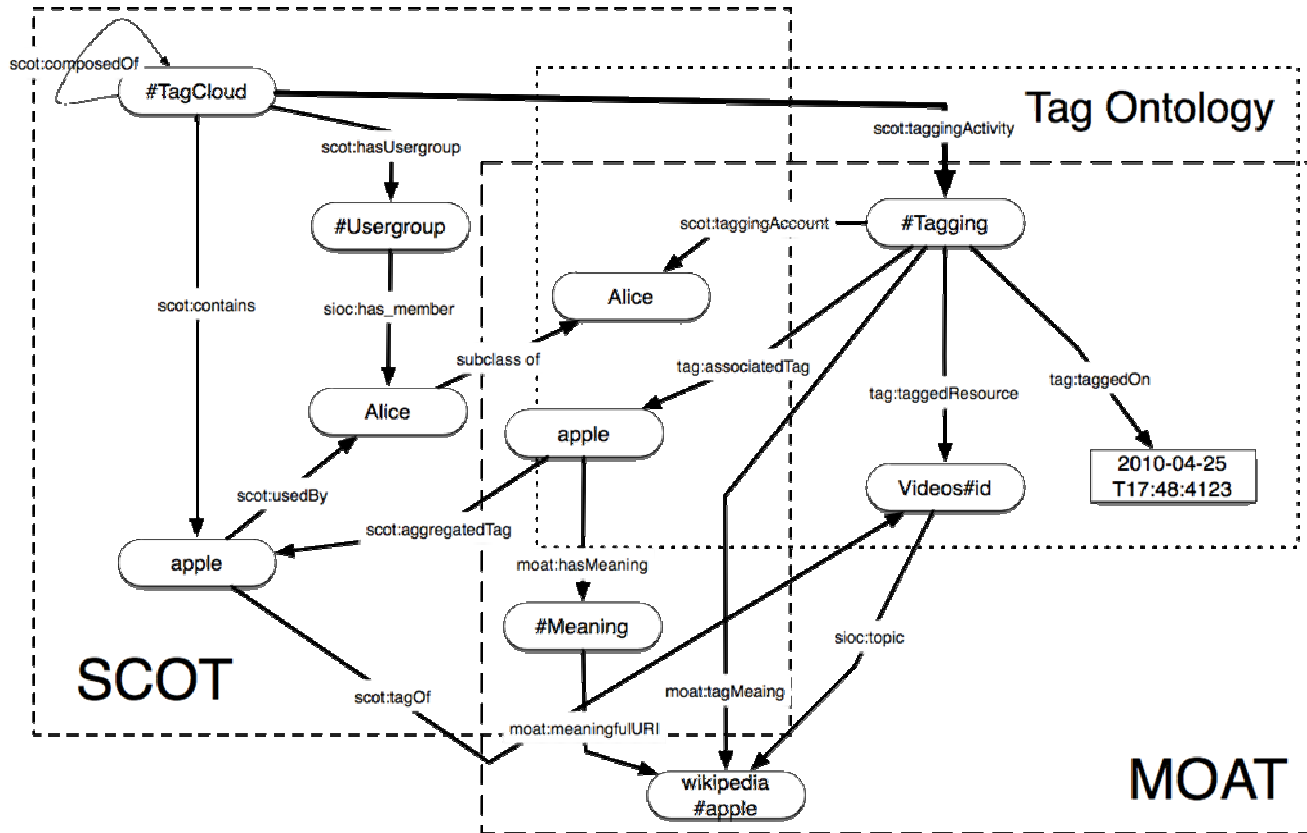


Figure 3 The combination amongst SCOT, Tag Ontology, and MOAT. Rounded rectangles refer to classes in tag ontologies and arrows refer to properties.

CONCLUSION

A wide variety of sites provide tagging features on many different types of resources and they also offer folksonomies which are popular as a collective way of categorising the resources. Nevertheless the concepts associating with social tagging are not well-defined. In this paper, we aimed to provide comprehensive overview of tagging and folksonomies by contrasting folksonomies and formal classifications, and then discussed the limitations of tagging such as ambiguities, sharing and exchanging tagging data. In the first half of this paper the relevant concepts are discussed with some of issues and challenges for improving current tagging systems. In particular, to achieve social ecology of tagging data we emphasise the consistent representation of tagging data, combined with Semantic Web technologies. Tag ontologies aim to describe tagging activities at a semantic level. By using tag ontologies it is possible to represent tagging data with appropriate semantics, and simultaneously to reuse and exchange the data across different sites. Several tag ontologies have been used in various websites. Since certain tag ontology cannot semantically represent overall tagging activities, there needs to interlink amongst those ontologies. The proposed approach in this paper offer improved opportunities for representing folksonomies with explicit tag meaning and the relationships of tagging entities. In the future, we will try not only to extend this approach into other RDF vocabularies but also to find out how this approach can be used in current tagging systems or folksonomies.

REFERENCES

1. Bailey, K. D. (1994), *Typologies and Taxonomies: An Introduction to Classification Techniques* (Quantitative Applications in the Social Sciences), Sage Publications, Inc.
2. Bizer, C., Cyganiak, R., Heath, T. (2007), *How to Publish Linked Data on the Web*, Available at: <http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTutorial/> (accessed 25 April 2010).

3. Berners-Lee, T. (2006), Linked Data, World Wide Web design issues, Available at: <http://www.w3.org/DesignIssues/LinkedData.html> (accessed 25 April 2010).
4. Fischer, D. H. (1998), From thesauri towards ontologies?, in Proceedings of the 5th ISKO Conference on Knowledge Organization.
5. Garshol, L. M. (2004), Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all, *Journal of Information Science*, Vol. 30 No. 4, pp. 378-391.
6. Gilchrist, A. (2003), Thesauri, taxonomies and ontologies – an etymological note, *Journal of Documentation*, Vol. 59 No. 1, pp. 7-18.
7. Golder, S. A. & Huberman, B. A. (2006), The Structure of Collaborative Tagging Systems, *Journal of Information Science*, Vol. 32 No. 2, pp. 198-208.
8. Gruber, T. (2008), Collective knowledge systems: Where the Social Web meets the Semantic Web, *Journal of Web Semantics*, Vol. 6 No. 1, pp. 4-13.
9. Gruber, T. (2007), Ontology of Folksonomy: A Mash-up of Apples and Oranges, *International Journal on Semantic Web & Information Systems*, Vol. 3 No. 2, pp. 1-11.
10. Gruber, T. (2005), TagOntology - a way to agree on the semantics of tagging data, Presentation to Tag Camp(www.tagcamp.org), available at: <http://tomgruber.org/writing/tagontology-tagcamp-talk.pdf> (accessed 25 April 2010).
11. Gruber, T. R. (1995), Toward principles for the design of ontologies used for knowledge sharing, *International Journal of Human-Computer Studies*, Vol. 43 No. 5-6, pp. 907-928.
12. Gruber, T. R. (1993), A translation approach to portable ontology specifications, *Knowledge Acquisition*, Vol. 5 No. 2, pp. 199-220.
13. Halpin, H., Robu, V. & Shepherd, H. (2007), The complex dynamics of collaborative tagging, in WWW 07: Proceedings of the 16th international conference on World Wide Web, ACM, New York, NY, USA, pp. 211-220.
14. Jacob, E. K. (2004), Classification and categorization: a difference that makes a difference, *Library Trends*, Vol. 52 No. 3, pp. 515-540.
15. Kim, H., Passant, A., Breslin, J. G., Scerri, S. & Decker, S. (2008a), Review and Alignment of Tag Ontologies for Semantically-Linked Data in Collaborative Tagging Spaces, in 2008 IEEE International Conference on Semantic Computing, pp.315-322.
16. Kim, H., Scerri, S., Breslin, J. G., Decker, S. & Kim, H. (2008b), The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies, in The 8th International Conference on Dublin Core and Metadata Applications, Berlin, Germany.
17. Kim, H., Decker, S. and Breslin, J. G., (2010), Representing and sharing folksonomies with semantics, *Journal of Information Science*, Vol. 36, pp. 57-72.
18. Knerr, T. (2006), Tagging Ontology - Towards a Common Ontology for Folksonomies, available at: <http://code.google.com/p/tagont/> (accessed 25 April 2010).
19. Kroski, E. (2005), The Hive Mind: Folksonomies and User-Based Tagging, available at: <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/> (accessed 25 April 2010).
20. 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, New York, NY, USA, pp. 31-40.
21. Mathes, A. (2004), Folksonomies - Cooperative Classification and Communication Through Shared Metadata, available at: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html> (accessed 25 April 2010).
22. Mejias, U. A. (2005), Tag Literacy, available at: http://ideant.typepad.com/ideant/2005/04/tag_literacy.html (accessed 25 April 2010).
23. Merholz, P. (2004), Metadata for the Masses, available at: <http://adaptivepath.com/publications/essays/archives/000361.php> (accessed 25 April 2010).

24. Neal, D. (2007), Introduction Folksonomies and Image Tagging: Seeing the Future?, The Bulletin of the American Society for Information Science and Technology, available at: <http://www.asis.org/Bulletin/Oct-07/neal.html> (accessed 25 April 2010).
25. Newman, R., Ayers, D. & Russell, S. (2005), Tag Ontology, available at: <http://www.holygoat.co.uk/owl/redwood/0.1/tags/> (accessed 25 April 2010).
26. Noy, N. F. & McGuinness, D. L. (2001), Ontology Development 101: A Guide to Creating Your First Ontology, available at: <http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html> (accessed 25 April 2010).
27. Passant, A., Laublet, P., Breslin, J.G. and Decker, S. (2009), A URI is worth a thousand tags: From tagging to Linked Data with MOAT, International Journal on Semantic Web and Information Systems, Vol. 5 No. 3, Special Issue on Linked Data.
28. Quintarelli, E. (2005), Folksonomies: power to the people, In the ISKO Italy UniMIB meeting, Milan, available at: <http://www.iskoi.org/doc/folksonomies.htm> (accessed 25 April 2010).
29. Quintarelli, E., Rosati, L., & Resmini, A. (2006). Facetag: integrating bottom-up and top-down classification in a social tagging system, in the Euro IA Conference, 2006, Berlin. Available at: <http://www.facetag.org/download/facetag.pdf> (accessed 25 April 2010).
30. Shirky, C. (2005), Ontology is Overrated: Categories, Links and Tags, available at: http://shirky.com/writings/ontology_overrated.html (accessed 25 April 2010).
31. Smith, G. (2004), Folksonomy: Social Classification, available at: http://atomiq.org/archives/2004/08/folksonomy_social_classification.html (accessed 25 April 2010).
32. Sowa, J. F. (1999), Knowledge Representation: Logical, Philosophical, and Computational Foundations, Course Technology.
33. Spivack, N. (2005), Folktoologies - Beyond the Folksonomy vs. Ontology Distinction, available at: http://novaspivack.typepad.com/nova_spivacks_weblog/2005/01/whats_after_fol.html (accessed 25 April 2010).
34. TagCommons Project (2007), Ontologies Vs. Formats Vs. Schema Vs. APIs, available at: <http://tagcommons.org/2007/03/02/ontologies-vs-formats-vs-schema-vs-apis/> (accessed 25 April 2010).
35. Vander Wal, T. (2005), Folksonomy Definition and Wikipedia, available at: <http://www.vanderwal.net/random/entrysel.php?blog=1750>. (accessed 25 April 2010).
36. Weinberger, D. (2005), Taxonomies and Tags: From Trees to Piles of Leaves, Esther Dysons Monthly Report, Vol. 23 No. 2, pp. 1-36.