JIS

Representing and sharing folksonomies with semantics

Hak-Lae Kim and Stefan Decker

Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland

John G. Breslin

School of Engineering and Informatics, National University of Ireland, Galway, Ireland

Abstract.

Websites that provide content creation and sharing features have become quite popular recently. These sites allow users to categorize and browse content using 'tags' or free-text keyword topics. Since users contribute and tag social media content across a variety of social web platforms, creating *new* knowledge from distributed tag data has become a matter of performing various tasks, including publishing, aggregating, integrating, and republishing tag data. However, there are a number of issues in relation to data sharing and interoperability when processing tag data across heterogeneous tagging platforms. In this paper we introduce a semantic tag model that aims to explicitly offer the necessary structure, semantics and relationships between tags. This approach provides an improved opportunity for representing tag data in the form of reusable constructs at a semantic level. We also demonstrate a prototype that consumes and makes use of shared tag metadata across heterogeneous sources.

Keywords: folksonomy; Semantic Web; social tagging; tag ontology

1.   Introduction

Web 2.0 sites that allow the creation of content by users or communities have become popular in recent years. The content of these sites can be virtually anything: blog entries, message board posts, videos, audio, images, wiki pages, user profiles, bookmarks, events, etc. Links on Web 2.0 can be created not only between people, but also between people and content, or among content items. Each content item can often be viewed as a social object that indicates why people affiliate with others or participate in communities [1]. According to Engeström [2], this object can enhance social relationships among participants who have an underlying shared interest within a community. Knorr Cetina [3] says that *object-centred sociality* refers to individuals grouped together around a shared object that mediates the ties between them. Therefore, we can assume that each content item on a Web 2.0 site can be a source of social connectivity, catalysing social networking in virtual spaces. However, creating new knowledge or new social relationships from various social objects remains

Correspondence to: Hak-Lae Kim, Digital Enterprise Research Institute, National University of Ireland, Galway, IDA Business Park, Newcastle Road, Galway, Ireland. Email: haklae.kim@deri.org

Hak-Lae Kim, Stefan Decker and John G. Breslin

a big challenge in terms of data sharing and interoperation. Since a content item is a heterogeneous resource, with many different associated type(s), and will be defined by different sets of metadata on distributed sites, a consistent way of collecting shared interests from independent and heterogeneous sites is required.

A key feature of user-contributed content in Web 2.0 sites is that the content item may be tagged, and can be shared with and commented upon by others. Tagging itself is a simple process, tapping into an existing cognitive process without adding much cognitive cost [4]. Tagging does not aim to create a strict classification of objects, but rather allows a user to categorize an object according to their own interests with their own keywords. Although a few words alone cannot identify user interests, a culture of mass participation leads to social interaction among users, and influences the use of terms in a community. This is the basis on which social tagging works. For example, tagging systems or 'folksonomies', such as on Flickr (www.flickr.com, a photo sharing site) or Delicious (http://delicious.com, a bookmark sharing site), enable users to categorize information sources in an unstructured way, and these sites often present popular tag usage visually via the use of *tag clouds*. Since users can browse content using the tags displayed in tag clouds, a tag is not just a keyword but also acts like a category for the associated content [4]. Tagging is not only a common feature of social content but also an important object for mediating common interests across independent, heterogeneous sources (e.g. tags on blog posts on the blogosphere). From this perspective, tags can be seen as objects for sharing, exchanging, and integrating a user's interests through tags attached to social objects on various Web 2.0 sites. Tag sharing can be an alternative method towards creating new knowledge from heterogeneous platforms. However, there are various issues related to interoperability in terms of tag sharing. Currently, conventional tagging systems lack the functionality for tag reuse between different platforms [5]; tagging activities arising from user participation are locked into host sites, meaning that tag data cannot be used to allow related users to be discovered and connected across heterogeneous platforms [6]. This may prevent the formation of new, additional connections between people who may have common interests.

To solve these limitations, we will introduce a novel approach that allows people to reuse and exchange tag metadata between each other and across different platforms. A semantic tag model for representing folksonomies is discussed and is followed by the introduction of our tag-sharing platform prototype. The remainder of this paper is structured as follows. In Section 2, we start by giving an overview of the motivations for this work. Section 3 provides a description of social tagging and folksonomies, and introduces the Semantic Web, while Section 4 presents several models for describing tagging activities and folksonomies. A brief overview of Social Semantic Cloud of Tags (SCOT) is presented in Section 5. In Section 6, we describe the int.ere.st web site, which aims to allow people to share semantic tag metadata. Section 7 discusses some issues that emerge in the process of sharing tagged resources. Finally, in Section 8 we present some of our conclusions and explore future directions.

2. Motivations

Since social websites began allowing users to tag their resources, users acquired a means to declare their interest in a wide range of topics as expressed through their tags, and they were also able to share their tagging practices with others within the same platform. In this context, a tag that is the result of user participation can be viewed as a social object, making it possible to mediate and interlink users across heterogeneous tagging platforms. As illustrated in Figure 1, if users have common tagging practices on Delicious and Flickr, they might have similar interests, and it would be easy to interlink them using similar tags. Then, they can share their resources, even though the resources may be of different types and are located on different platforms. If we could expand this scenario, a linked tag space could be constructed by interlinking both centralized and decentralized tagging datasets. Isolated tags in conventional tagging systems can become intertwined, be discoverable and become integrated through emergent mash-up technologies. This provides an additional opportunity that leverages social connections for knowledge sharing. As social tagging

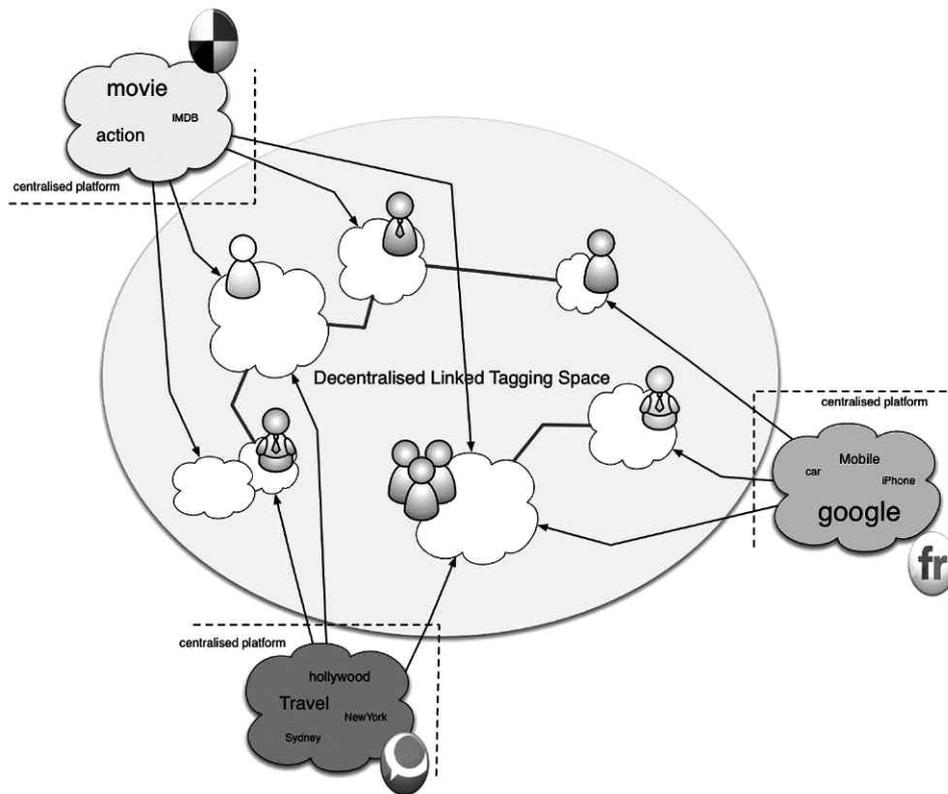Hak-Lae Kim, Stefan Decker and John G. Breslin

Fig. 1. Linked tag space.

implies sharing and interaction between users via tags, individual users can share their tag data and can extend their networks based on similar patterns of specific tag usage. Although these links might not reflect strong relationships between users, the idea supports the 'strength of weak ties' notion popularized by Mark Granovetter [7]. A weak tie is a connection that is relatively weaker than the 'regular' connections of a node (e.g. close family versus acquaintances). Granovetter emphasizes the importance of weak ties in information flow and explains how these ties have a special role in bridging or mediating information between nodes. In a linked tagging space, the tagging practices of users play a role in improving tag-centred sociality as weak ties.

To achieve these goals, we take into account some issues, including the need for a standard representation of tags and interoperability mechanisms for sharing tags across platforms. In particular, we will focus on the following issues:

- Common conceptualization: since tag data are ill-structured and lacking in semantic control, a consistent way of exposing and accessing tag data is a key aspect for supporting interoperation among heterogeneous sources. An ontology can provide the appropriate semantics to describe tagging entities and the relationships between them.

- Reuse tags and folksonomies: although existing tag ontologies offer a consistent representation for tags, they are somewhat limited when it comes to expressing various aspects of folksonomies. In order to enable the reuse of tags across platforms, some alternative technologies, such as the Semantic Web, can be considered to more efficiently achieve these scenarios.

- Linked tag data: users may have a number of accounts on the web. If users have similar interests, or assign similar tags to different resources, it might be useful to make links between these different sets of tag data. There are many opportunities for building linked datasets if tagging practices can be linked together.

3.   Literature review

3.1.   *Tagging and folksonomies*

Tagging is an approach for metadata creation in which users describe or organize information objects with a free-form set of keywords (tags) [4, 8, 9]. There are many definitions for the various activities involved in tagging, but the main constituents are:

- Tag: a word or phrase for a resource that has meaning to a human being, and is not taken from any predefined and controlled terminologies.

- Tagging: the process of adding tags to a resource, without the constraint of a controlled vocabulary.

- Folksonomy: an aggregation of tags for multiple resources, shared by multiple users.

In general, people can use any term as a tag without exactly understanding the semantics of the chosen terms. The tags produced by a number of users can be aggregated to form a non-hierarchical taxonomy [8]: a *folksonomy*, as coined by Vander Wal [10]. The collective and self-motivated approach of social tagging offers novel opportunities to users, including flexibility with information organization, enhanced findability, and serendipitous browsing with respect to information activities [4, 9]. The ease of participation leads users to express their interests in diverse resources to create a bottom-up consensus view of the world [11]. This massive participation of users leads to emergent semantics for shared interests via the process of aggregation. However, the simplicity and ease of use of tagging creates a number of limitations during tagging activities. Ambiguity of tags is one of the problems inherent in an uncontrolled vocabulary and the lack of synonym control can lead to different tags being used for the same concept [4].

Methods for researching the contributions of social tagging and folksonomies have been carried out. Golder and Huberman [12] analysed the structure and usage patterns of tagging systems and discussed the distinction between collaborative tagging and taxonomies. Marlow et al. [13] focused on the social aspect of tagging and proposed a model where 'tags are represented as typed edges connecting users and resources'. Halpin et al. [14] analysed the dynamics of a collaborative tagging system by focusing on the 'short head' rather than the 'long tail', combined with measures of stability of tag frequencies and information value (the measure of a tag based on the number of pages it retrieves). They also extended a tripartite model for tagging [15], which consisted of users, tags, and resources, using a preferential attachment model. Cattuto et al. [16] highlighted the network structure of Bibsonomy and Delicious and also find similar characteristics to a small-world network. Furthermore, they looked at relative path lengths across the tripartite network, and studied tag co-occurrence, noting that 'evidence is compatible with the existence of complex, possibly hierarchical structures in the network of tag co-occurrence'. These hierarchical structures are the subject of ontological deduction research. In addition to the semantic issues of tags, tagging systems themselves only give users restricted functionality for reusing, sharing and discovering tags [5]. To date, however, relatively few studies have been devoted to such issues.

3.2.   *Tagging on the Semantic Web*

The Semantic Web, consisting of a 'web of data', enables machines on the web to understand the requests of specific users and allows these machines to use web content semantically and automatically [17]. It aims to provide a common framework that allows data to be shared and reused across applications, enterprises and communities [18]. Semantic Web-based approaches can support a standardized metadata schema to represent both the structures and the semantics of tagging data.

Tagging is not only a way for representing concepts through cognitive association by individual users, but it is also a social and democratic process that encourages social relationships among users [19]. Thus, tagging on the Semantic Web can represent the overall features of tagging entities in a given community or site, while simultaneously allowing for a continuous transformation from individual to social tagging with appropriate semantics. Gruber [20] emphasized the need for folksonomies and

ontologies to work together, and tried to identify and formalize a conceptualization of tagging data at a semantic level. Typical social tagging systems do not provide explicit links between the involved entities, nor do they expose their data in a standard form. The design of the *tag ontology* was an attempt to provide a common conceptualization of 'what tagging means' by providing a standardized way to collect, interpret and use tagging data [20]. One of the advantages of this ontology is that isolated tagging data can be easily made mobile and can be integrated across applications. Tags, users and their relationships in particular applications can be represented in the form of ontology using RDF or OWL, and these data can be exposed, shared and connected to other data via dereferenceable URIs on the web as Linked Data [21]. The term 'Linked Data' was coined by Berners-Lee to address the idea that data on the web can be connected to other data by meaningful hyperlinks [22]. Combining Linked Data with tagging practices can be viewed as a starting point for sharing and exchanging separate tagging activities on different platforms.

Most agree on what the most elemental building blocks of a tag model should be [23, 24]. Gruber proposes an ontology for folksonomies that can act as an infrastructure to build an ecosystem of tag data sources, services, agents and tools [20]. In his model, the core concept is *Tagging*, i.e. the act of associating tags with an object or item [20, 25]. Tagging comprises the following core concepts: a *Tag* is a word or phrase that is recognizable by people and computers; an *Object* is a thing to be tagged, identifiable by a URI or a similar naming service; the *Tagger* is the agent doing the tagging, such as the user of an application; the *Source* is the scope of namespaces or a universe of quantification for the object; and the *Polarity* is a vote for or against the assertion of the tagging. This model can be considered as a first step towards a generally-applicable representation model for tagging (even though it is not itself represented using Semantic Web vocabularies), as it clearly reveals a generic conceptualization of tagging. Newman's model [26] describes the relationships between an agent, an arbitrary resource, and one or more tags. This model consists of three core concepts: *Tagger*, *Tagging* and *Tag*, which represent a tagging activity. The concepts are serialized in RDF/OWL. MOAT (Meaning of a Tag)[1] is intended for semantic annotation of content by providing meanings for free-text tagging [27]. In addition to the extensions to the core concepts from Newman's ontology, MOAT provides the meaning class to represent customized, user-provided 'meanings' for tags. This class allows the meaning of tags to be unambiguous. The Nepomuk[2] Annotation Ontology (NAO) [28] is provided for annotating resources on the Social Semantic Desktop [29]. It is not entirely dedicated to tagging practices but, nevertheless, it demonstrates the increasing importance of tagging representations on social systems. These models, as a form of common agreement for tags, offer consistent structures, semantics and machine processability for tagging data, but they still do not support interoperability and collaborative cross-application tag sharing.

4. Models of social tagging

Before addressing the proposed tagging model, some details regarding the issues mentioned in Section 3 are discussed along with conceptual models. There is agreement that tagging entities include users, tags and resources. This is often called a *tripartite* model: **Tagging (U, T, R)**. **U** refers to the set of taggers that participate in the tagging activity, **T** identifies the set of tags that are assigned to resources, and **R** is the set of resources being tagged. This model has been considered as a generally applicable representation model for tagging. Existing tag ontologies, as noted earlier, are based on, or are an extension to, this model. Cattuto et al. [16] elaborated on the formal definition of this model with a user-based assignment of tags to resources: **(U, T, R, Y)** where **Y** is a ternary relation between entities (i.e. $Y \subseteq U \times T \times R$). In current tagging systems, tags are linked to all resources where a tag is assigned and these tags are also connected to users. However, since this information is implicitly stored in a database, an explicit representation method is required to describe all of the relevant information for the tagging instances. For example, if a tag is added when a content item is made, the tag creation data are normally assumed to be the same as the content item creation date (although it is not stored as this) and can be explicitly defined in this way. Figure 2(a) illustrates the relationships between entities with no semantic relationships, while the relationships between them, shown
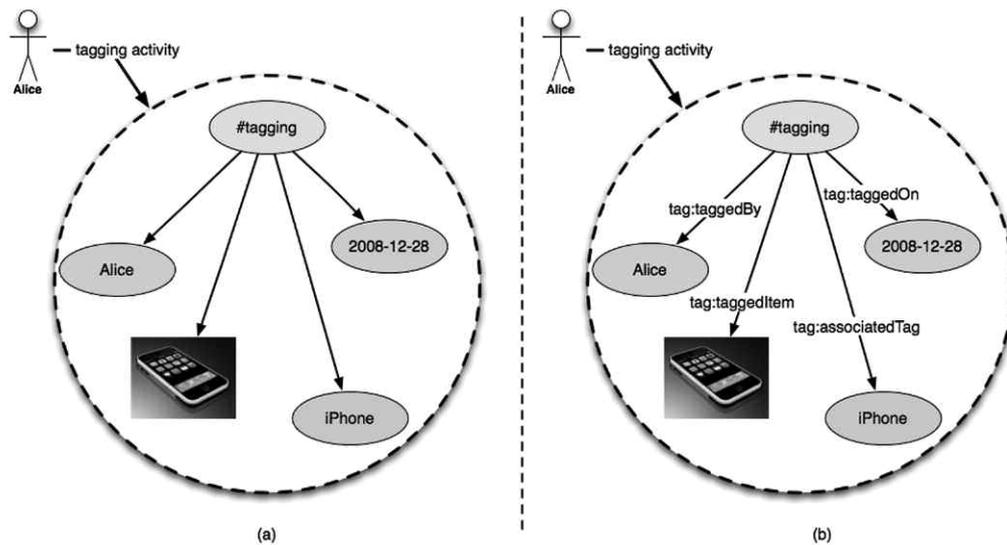
Fig. 2.   Representation of tagging entities and their relationships.

in Figure 2(b), are described in an explicit way using ontological representations. This shows in brief why Semantic Web technologies can augment social tagging. Some properties to deal with the inherent features of tags, however, are also needed.

4.1.   Semantic enrichment of tags

A tag label can have different variations, such as capitalization, singular versus plural, or a tag with special characters. In Figure 2, the tag *iPhone* can be written using many different conventions in the real world: *IPHONE* in upper case, *iphone* in lower case, or *i-phone* as a compound term. All variations have the same meaning, if not the intended purpose. Moreover, people can conclude that the tag *apple* refers to the computer company rather than the fruit because of its context – if the tag is used with the label *iPhone.* In addition to the variation issues, we need to consider semantic issues associated with tags, such as polysemy (i.e. the same term can refer to different concepts), synonymy (the same concept can be defined using different terms), and different lexical forms (different noun forms, different languages, acronyms) arising when using tags.

Moreover, rather than focusing on expressing a single tagging instance, we can examine the representation of the collective aspect of tagging practices, since folksonomies are the result of aggregating all tagging instances. The relationships between tagging entities should be dynamically changed depending on the context of the tagging activities: how many users are involved, how many tags are used or where the tagging occurred. Suppose that Alice creates different tagging instances over time and uses multiple tags (e.g. 'iphone' and 'apple') to categorize the resources. If the tag *iphone* is used in both tagging instances, the popularity of this tag and the tag's date of use have to be updated each time. If she uses both tags together, a co-occurrence relationship is created. This numerical property can be a useful feature for representing these co-occurences and for detecting emergent semantics in folksonomies. In brief, each tag has a popularity score, being the sum of total occurrences. Thus, the tagging model can be extended as follows: ***Tagging (U, T, R, Y, ∠)*** where $\angle$ is a relationship between tags (i.e. $\angle \subseteq U \times T \times T$) that includes the tag's lexical and numerical properties.

Since existing tag ontologies tend to focus on representing tagging activities, more properties in the tag ontology are required to deal with the inherent features of tags in order to assign an explicit meaning to those tags. On the other hand, tag ontology needs to (be extended to) represent the relationships between tagging instances as well as the collective features of tagging activities. Existing tag ontologies, including Gruber and Newman's models, do not cover this issue, since they focus on expressing individual tagging instances.

4.2. From personomy to folksonomy

As increasing numbers of tags have been created on a variety of social sites, a specific problem has arisen. Users' tagging activities are carried out in distributed environments. However, although the users use common concepts to tag various resources across different platforms, in reality they are using different collections of tags on different platforms. Tagging activities can consist of diverse knowledge domains and terminologies. Since people can have different or similar sets of tags across different platforms and the relationships among the entities can also differ, their tags and the links they generate are dynamically contextualized according to their specific interests. When users want to construct their knowledge model based on their distributed contributions, the information being aggregated needs to be customized for the users' context. The desire to deliver user-centric tagging practices calls for the reuse and sharing of tagging activities across host sites.

This is often referred to as a *personomy*, which consists of a single user's tagging practices based on their interests and independent of any platform. A personomy has a specific focus on personas. Note that a personomy can be constructed by aggregating tags reflecting a user's intention or purpose (from heterogeneous platforms); whether the user participates in tagging activities or not. As Gruber points out [20], we need to make explicit some notion of source or tag space, which one can consider as the scope of a tagging activity. Formally, a personomy P can be described as: $P = (u, R, T, S)$, where u refers to a single user, and S refers to a set of spaces in which tagging activities occurred. From this perspective, a folksonomy is a collection of personomies placed on heterogeneous platforms without boundaries (i.e. $F = \sum P$). We refer to this as a *platform-independent folksonomy.* Since this type of folksonomy would enable interlinks between tagging entities across platforms, one can aggregate, share and explore tagging data through a decentralized data space. We note that, in order to express personomies and folksonomies, one needs to consider some issues, including representation and sharing aspects. First, the data maintained on a platform cannot be easily reused on another platform. Second, each platform uses different formats within its own boundaries. There is no standard that regulates how tags and relationships between tags, users and resources are represented.

In summary, tagging practices contributed by users are locked into their host sites, which means that users cannot expose their own knowledge from these sites despite their needs. Since users' tagging activities have been carried out on diverse platforms, aggregated tagging information needs to become interoperable across these platforms. Although tag ontologies can be a promising solution for this purpose, existing models need to be extended, not only to have a standard representation for tagging activities, but also to mix or integrate personomies in a standard format that fits semantically. Technical issues are also evident in terms of supporting the reuse and exchange of personomies or folksonomies. One can use open APIs, syndication or other standard methods with existing Semantic Web ontologies to export, mix or maintain tagging data from distributed tagging platforms, but services will have to transform data from a native store to an interoperable format, and vice versa.

5. SCOT: semantic model for folksonomy

5.1. Overview

SCOT (http://scot-project.org) is an acronym for 'Social Semantic Cloud of Tags'. The name was chosen to emphasize the goal of providing a consistent framework for expressing tagging activities in a machine-understandable way [19]. This ontology represents the main concepts and properties required to describe information for tagging activities on the Semantic Web: it offers a collection of basic terms to describe tagging entities and their relationships in an explicit way using RDF/OWL.

Both the scot:TagCloud and scot:Tag classes play a key role in the representation of social tagging. The class scot:TagCloud is a specific type of container for grouping metadata relevant to tagging practices. In practice, the term 'tag cloud' is a common method to visualize tags or folksonomies. In the context of SCOT, this term as a concept is used to explicitly describe social aspects among the entities derived from tagging activities. Thus, this class is a type of sioc:Container, in that it contains tag items.

Some of the properties of TagCloud include: when the tag cloud was generated (dcterms:created), where the tagging occurred (scot:tagSpace), how many tags (scot:totalTags) and posts (scot:totalItems) it has. In particular, scot:hasUsergroup represents a set of users who have contributed the tags contained within a particular group. The property scot:composedOf describes a part of a TagCloud: for example, if a TagCloud consists of more than two TagClouds, the property identifies each of them with the specific URI. These two properties are an essential feature for discriminating between personomies and folksonomies: if the property scot:hasUsergroup has a single user and the scot:composedOf has a pair of URIs, the TagCloud is a type of personomy.

As a subclass of tag:Tag from Newman's model, the class scot:Tag describes a natural-language concept, which is used to annotate a resource. The purpose of this class is to describe the semantics and collectiveness of tags that are aggregated from individual tagging activities. As we discussed, tags can be created using many different conventions in the real world. This leads to the creation of critical barriers for more precise categorization and more useful navigation. The tag class has some properties for eliminating tags' ambiguity:

- scot:spellingVariant refers to variations in the way in which a word is spelt;

- scot:delimited is used to describe a multiple-word tag name where each word is separated by a certain character;

- scot:synonym defines synonymous terms.

These properties can reduce tag ambiguity resulting from the use of different conventions and even recommend more common patterns for tag names. Furthermore, in order to represent tag frequencies, SCOT introduces two properties: scot:ownAFrequency and scot:ownRFrequency. The former is intended to describe the absolute value of popularity for a specific tag. The purpose of the latter is to represent the relative value in order to identify the significance of a tag proportional to all the tags. A single tag can have both frequency formats. The popularity of a tag plays a key role in distinguishing its significance in folksonomies.

All earlier tag ontologies including Gruber [20] and Newman's models [26] do not provide a way of fully representing the meaning of a tag and the relationships between tags, since they focus on expressing individual tagging instances. On the other hand, SCOT offers various properties for representing tag semantics and collective characteristics of tagging entities [24]. These features are a cornerstone towards being able to identify, formalize and exchange a common conceptualization of tagging activities at a semantic level. Table 1 summarizes the core concepts of existing tag ontologies with their supports.

Table 1
Representative features of existing tag ontologies

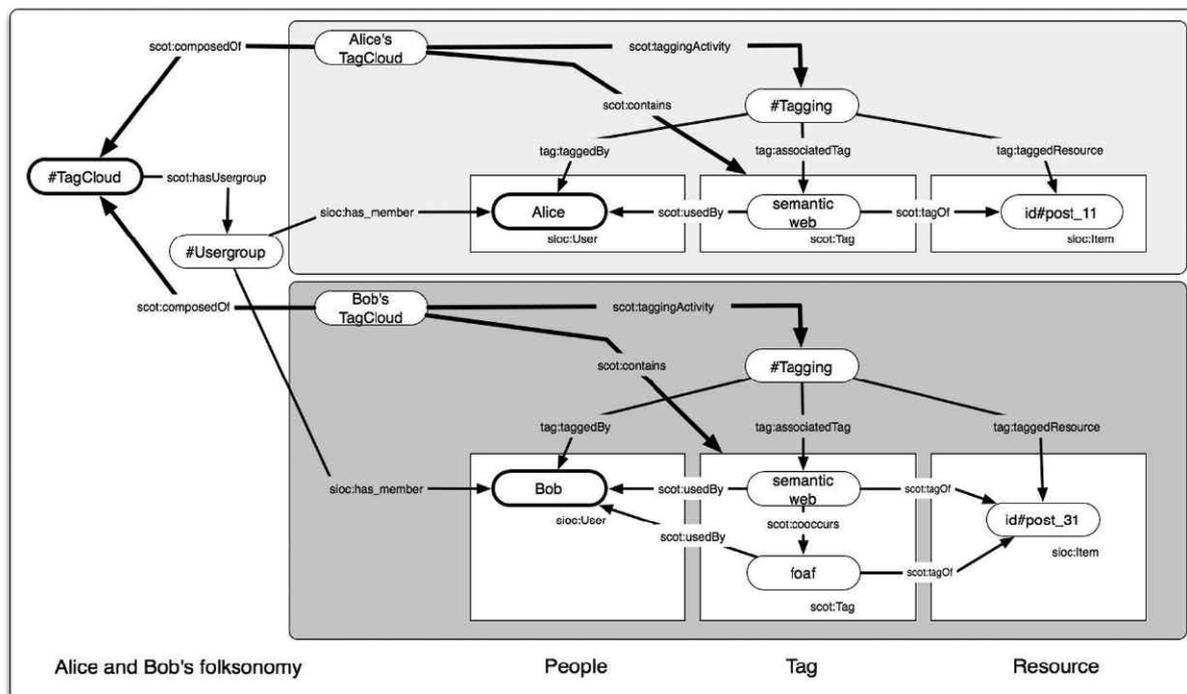| Ontology	Format	Tag	Tagging (Tripartite)	Tagger	Tag cloud	Tag meaning	Object's frequency
			Supports modelling for				
Gruber's model [20]	N/A	✓	✓	✓	✗	✗	✗
Newman's model [26]	OWL	✓	✓	✗	✗	✗	✗
MOAT	OWL	✓	✓	✗	✗	✓	✗
NAO	RDFS	✓	✗	✗	✗	✗	✗
Knerr's model [30]	OWL	✓	✓	✓	✗	✗	✗
Echarte's model [31]	OWL	✓	✓	✓	✗	✗	✗
SKOS Core	OWL	✓	✗	✗	✗	✗	✗
SIOC	RDFS	✓	✗	✓	✗	✗	✗
Annotea	RDFS	✓	✗	✗	✗	✗	✗
SCOT	OWL	✓	✓	✗	✓	✗	✓

Fig. 3. Platform-independent folksonomy by combining Alice and Bob's personomies.

5.2. Expressing personomies and folksonomies with semantics

Now we will show how to express personomies and folksonomies at a semantic level using SCOT. Figure 3 shows the merged folksonomy from both Alice and Bob's personal tag clouds (i.e. personomies). In this example, we assume that both users have their own tagging practices on different spaces, while they may also have similar tagging behaviours.

A tagging event is represented by tag:Tagging with relevant properties (i.e. tag:associatedTag, tag:taggedBy, tag:taggedOn and tag:taggedResource) in Newman's ontology. This class provides metadata related to tagging activities, connecting basic entities such as users, tags and resources. SCOT introduces some approaches to defining collective and aggregated properties of tagging activities. For example, the scot:taggingActivity describes a relationship between scot:TagCloud and tag:Tagging. Thus, all tagging events for a user are collectively linked to an instance of the TagCloud class. Multiple tags in tagging events are aggregated to one unique scot:Tag, if the names of the tags coincide. At the same time, occurrences of the tags are updated via two properties: scot:ownAFrequency and scot:ownRFrequency. The scot:lastUsed and tag:taggedOn properties provide for metadata related to time. scot:Tag is linked to scot:TagCloud via the scot:contains property. Tagger (user) information is represented using SIOC. SCOT also introduces the scot:taggingAccount property, describing the relationship between a tagging activity and the account used when performing the tagging.

To describe collective tagging activities for multiple users or communities, SCOT introduces the scot:composedOf property to link multiple TagClouds. With this property, Alice and Bob's personomies can be interlinked. The tagging information of both users can be stored within one tag cloud (i.e. folksonomy) and simultaneously interlinked between them. The SCOT project provides some applications to generate SCOT metadata and manage, share and search these data on the web.

This approach shows how a user-centric folksonomy for representing the interests of small groups or communities can be created in SCOT. The shared tag clouds allow the interlinking of multiple users and their personal tag clouds. By also taking into account relationships in between tags (e.g. co-occurrence), folksonomical data can also be enhanced. It is also possible to adopt this example across sites or resources. Although both users have tagged different resources, the entities involved in the tagging
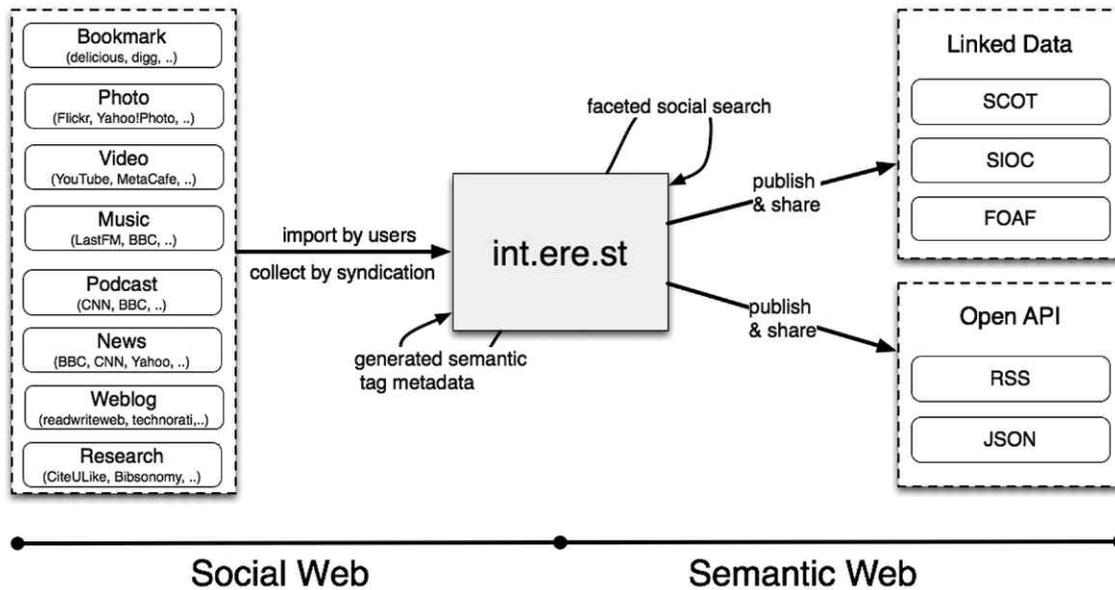
Fig. 4.   Functional architecture of int.ere.st: this site supports a set of standards, such as OpenID, RSS, and RDF, and offers some proprietary APIs to collect and share semantic tag metadata.

activities are explicitly linked to each other and the structure of the tagging data enables it to be shared and reused. Therefore, this approach can be adopted to create a customized folksonomy for those users. As with all RDF data, SCOT can be utilized by SPARQL [32], the query language for the Semantic Web.

6.   Let's share tags: int.ere.st website

int.ere.st is a prototype of a tag-sharing platform conceived for reusing tagged resources across heterogeneous platforms. A major goal of int.ere.st is to create a Semantic Web-based tagging application capable of solving the common problems of tags and tagging systems [19]. Throughout the implementation of this site, requirements have been made that aim to enrich the semantics of tagging entities and this allows us to easily manipulate semantic tag metadata. In other words, int.ere.st allows users to reuse, manipulate, and republish semantic tag metadata across heterogeneous platforms. Using machine-readable RDF metadata, this site is able to solve common problems related to tags, such as lack of structure and lack of semantics for tags. As a consequence, the combination of tags and Semantic Web technologies enables reusability through an agreed meaning of tags (thereby improving information retrieval via these tags).

From a technical point of view, int.ere.st is built on a variety of technologies: Apache, PHP, and MySQL. These frameworks are used to implement most functions in typical social websites. The need to provide interoperability between tagging information has led to the use of Semantic Web technologies – RDF and SPARQL. The majority of social websites now provide APIs based on popular mechanisms (e.g. REST, SOAP, and XML RPC). These APIs provide community users and applications with easy and intuitive access to data from the sites. Rather than storing syndicated information from host sites, int.ere.st transforms the data into the machine-readable RDF format using SCOT and SIOC. This functionality is achieved through the combination of Semantic Web and mash-up technologies. This site uses a tag cloud as the default visualization method for tagging data. Figure 4 shows some of the functionalities of int.ere.st. Users can manage a collection of their tagging data across different platforms. The tagging data are published in SCOT as Linked Data, while open APIs are also used to encourage the sharing and the exchange of tagging data. int.ere.st shows how user-contributed and user-distributed tags can be effectively intertwined with each other
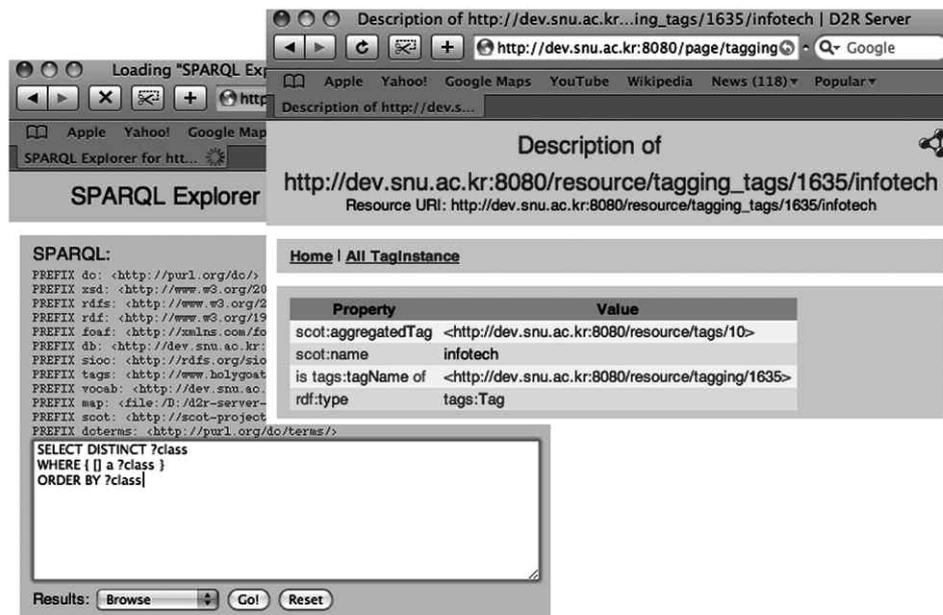
Fig. 5.   SPARQL endpoint for int.ere.st.

using richer machine-readable formats to improve the reusability of isolated tags from each platform. Throughout the process of tag sharing, users will have the opportunity to get a live, hands-on experience of the combined Social Web and Semantic Web features.

6.1.   Content aggregation and semantic enrichment

In int.ere.st, a number of tagged resources are automatically aggregated from a variety of applications, services and sites, and users can also import their data into the system after registering. Once a source is collected, a user can check for new content at user-determined intervals, and retrieve the updated information. Therefore, a user's content items are regularly checked for updates. Through this task, tagging data are lifted out of the 'walled gardens' in which it resides to a more open architecture. Then, the common conceptualization for representing tagging data and the mapping between RDF vocabularies is leveraged. The aggregation and representation follows the principle that each source with a unique URI is transformed into a SCOT instance.

The linking of tagging entities can be done with SCOT and queries for particular information can be made using SPARQL. As the amount of aggregated sources may increase exponentially, it is difficult to generate and update semantic tag metadata synchronously. To solve this problem, we use the D2R Server [33], a tool that maps relational databases to RDF and that is accessible through SPARQL. By using this tool, semantic tag metadata in the system exposes the time when a request is made. Furthermore, the site supports an HTTP content negotiation mechanism to provide users with either HTML or RDF data representation and a SPARQL interface, which allows users to query tag metadata directly.

6.2.   Social search and browsing

There are several ways to search tagging entities on the site. First, a tag search allows users to look for similar tagging patterns or for persons who share their interests. The basic search operators are: *and*, *or*, *co-occurring tags*, and *broader and narrow tags*. The operators enable users to restrict their search conditions. Each operator is translated into a SPARQL-based query format at run time. A simple type of query is based on a combination of tags, such as co-occurring tags, while int.ere.st allows

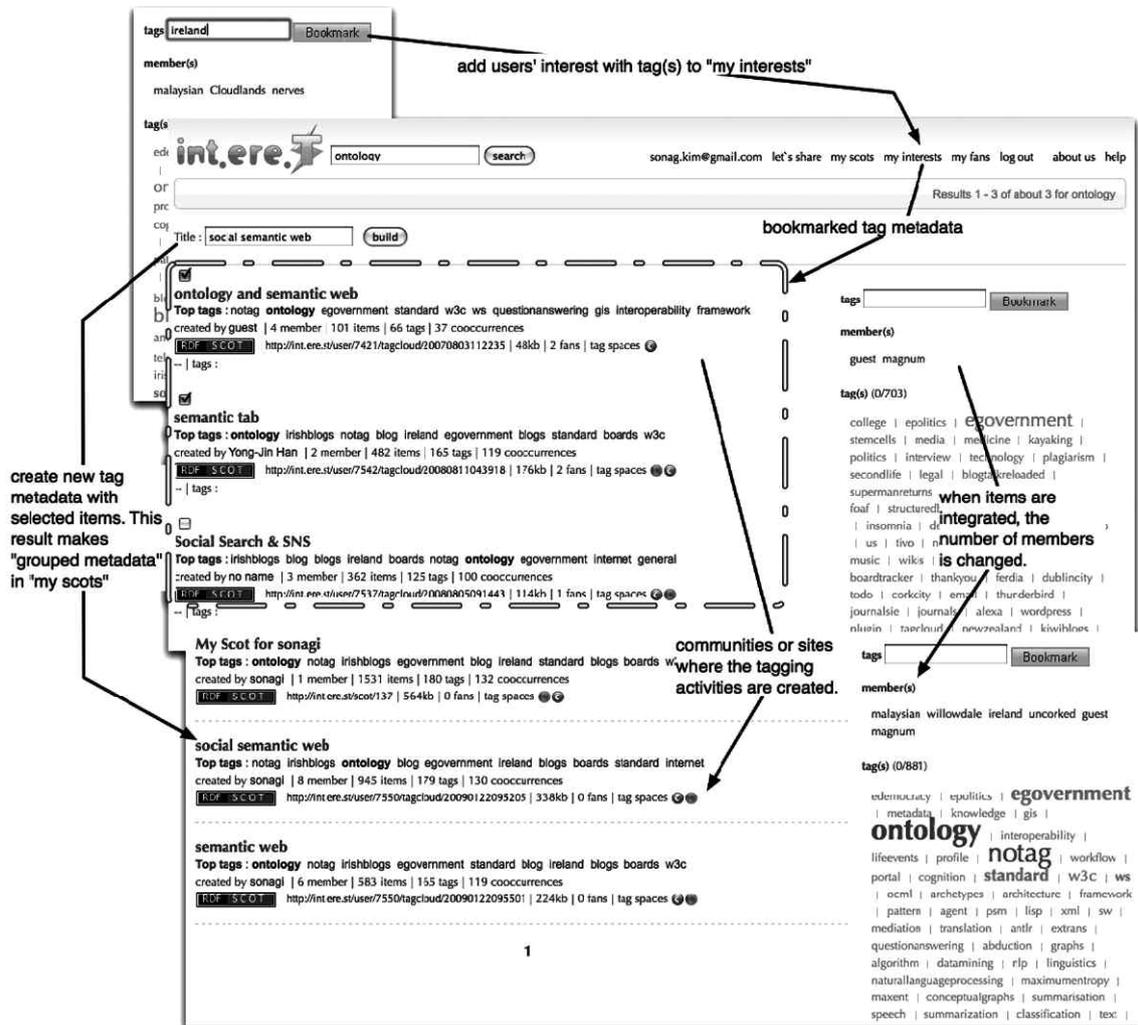Hak-Lae Kim, Stefan Decker and John G. Breslin

Fig. 6. A user-generated platform-independent folksonomy. A user can bookmark and combine the tag metadata from other individuals to add it to their own interests.

users to operate with more sophisticated queries, e.g. '*return a user using a particular tag in PlanetRDF and CiteULike*' or '*show users with a set of tags in CiteULike*'. Also, a simple social browsing system is supported. When *created by* from search results is clicked, all metadata instances created by the individual or community are listed.

6.3. Sharing interests with semantic vocabularies

If a number of users have similar tagging practices, it may be possible to interlink them across heterogeneous platforms. In order to achieve this scenario, tagging data collected from distributed sites need to be mapped or integrated. A SCOT instance can be classified into several types: imported, bookmarked and grouped. The imported type refers to a metadata instance from different external applications. For example, an instance generated by the SCOT Exporter [34] can be imported into the system. The bookmarked type refers to instances for which users create links when they want to remember and/or share them. The grouped type is an integrated metadata set comprising several bookmarked instances. The bookmarked type is created by other users, while the grouped type is created by a logged-in user. Imported metadata can be one of the two types mentioned above. This is also the result of users' contributions to the system.

Different types of metadata instances can be used to make a social connection between users based on their shared interests. There are two approaches: a *positive* method and a *passive* method. The *positive* method refers to the fact that users can create a network by bookmarking or by building grouped instances. For example, if users are interested in a certain instance, including tagged resources, they can create a link with tags as a bookmark, which is saved in the *my interest* view of the system. They are also able to integrate some bookmarked instances according to their specific purposes or interests. A list of grouped instances is located in the *my scot* view. The links in the instances generate networks between the users. Meanwhile, the *passive* method refers to the bookmarking of people's activities. The *my fans* view shows users of this approach. The number of fans shows how many users refer to a specific user, or the number of users to which the user is referring to. In this process, social connections are based on the users' participation and tagged data play a key role in building these user-contributed networks. Figure 6 illustrates the overall sharing process for tag metadata.

On the website, SCOT exposes tagging data consistently, while using SIOC [35] and FOAF [36] to represent other relevant information in a machine-processable way. For example, a personal profile is represented using FOAF and SCOT in the system, and grouped types in the system are mapped to SIOC. All bookmarked instances are represented by the foaf:interest property in FOAF, and all types of SCOT instances for a certain user are mapped to sioc:Items. This process can be achieved automatically. This interlinking mechanism among the dedicated vocabularies provides a way to enhance interoperability among independent applications. Given the reality of heterogeneous tagging systems, this approach provides a standardized method to compare, combine and manipulate tagging data. Using Semantic Web technologies to mash-up public APIs can provide a standard access method for tagging data, and it also gives the potential to link tagging data because such RDF vocabularies provide methods for interlinking different sources (people, concepts, groups, etc.) at a semantic level.

7.   Discussion

We must take into account many different issues regarding the process of tag sharing across different platforms. In particular, we will focus on copyright and social network issues.

The majority of people participating in content creation and dissemination are not trained in legal matters and may not give due consideration to copyright and legal issues. As a growing number of users participate in content sharing, the risk of infringements increases. There are some collaborative efforts in terms of licensing and copyright issues for open data [37]: Open Data Commons (www.opendatacommons.org), ccREL [38], ODRL (Open Digital Rights Language) [39], and IPROnto [40], etc. These approaches help one to describe copyrighted information in an explicit manner. Copyright and intellectual property are normally related to content rather than tags. However, since tags are essentially assigned to a particular resource or content item, the tag sharing process implies that the associated content and the metadata (e.g. tags) will be shared together. It is for this reason that copyright issues need to be considered in the process of tag sharing. We propose a way to interlink existing copyright ontologies and SCOT. In the context of SCOT, current research efforts have been directed toward the introduction of the scot:hasCopyright and scot:hasIPR properties. The scot:hasCopyright property identifies copyrights, access restrictions, manipulation limits and republishing guidelines for a given resource. The property scot:hasIPR identifies and groups a set of copyright information such as DRM (Digital Rights Management), terms and conditions, services and usage restrictions. More importantly, both properties are designed to cater for the inclusion of any existing and/or future copyrights that contain the URI of the licence applied through rdfs:Resource range values for their properties. When copyright is described by an ontology, it is possible to link to a specific item, whereas copyright information such as 'terms of conditions' may be linked to an associated URL, which includes a detailed description. Copyright for a set of resources, such as that within a TagCloud, can be described with the scot:hasIPR property. It is important to note that the proposed approach does intend to link a tagged resource to appropriate license or copyright URIs on the web, rather than assigning copyright information to the tags themselves.

From a social network perspective, SCOT is limited and is not intended to fully represent user profiles. Thus, we propose an extended solution that interlinks SCOT with existing RDF vocabularies to address other vocabularies' limitations with respect to representing tagging data. FOAF enables users to describe a set of tags in their profiles via skos:Concept. This method, however, does not enable users to link to multiple sets of tags generated through heterogeneous sources. For example, if a user participates in both Delicious and Flickr, it is not easy to link to their tagging data. Tagging data can potentially be included as part of a user's FOAF profile. This is possible through SCOT by linking a SCOT instance (i.e. a set of tags) via foaf:interest to indicate an 'interest' of a foaf:Agent. This method also supports data portability since decentralized tagging data could in this way be re-used. We propose the development of recommended models to describe how tagging activities can be connected to people, for example, by using a scot:tagCloud property to connect a TagCloud to the foaf:Person or sioc:User who created it.

8. Conclusion

This paper described several components for reusing one's personal set of tags, including the semantic model designed to meet requirements, and various export and sharing methods. Defining the relationships between tags is one of the benefits of using an ontology to model folksonomies. This can be viewed as a means of adding appropriate semantics to tagging entities. Tag ontologies, using the universal RDF data model, play a key role in realizing data integration and interoperability across platforms. The SCOT ontology provides a formal representation of tagging data to express, share and integrate these data and inherent semantics across platforms. This ontology has a number of advanced tagging properties that can be used to detail the popularity and usage of tags, along with deduced relationships between tags. We have introduced a decentralized tag sharing approach, allowing users to reuse shared tag resources. int.ere.st is the first open-tagging platform for the Semantic Web that aims to make tagging data open, more universal, and available for applications across social tagging sites. In order to allow users and developers to support social capabilities underlying these tagging data, the platform also provides some open APIs. In addition, the tagging data are published in the SCOT format as Linked Data, while open APIs encourage the sharing and exchange of tagging data. In the future, we plan to interlink our data to the Linked Data cloud, such as DBPedia, and Revyu.com. Through this, a social search based on tags will be able to offer a more powerful and meaningful service to users. This approach will benefit online communities through the provision of a seamless framework for accessing and interlinking social content across different applications.

Endnotes

1 See http://moat-project.org
2 See www.nepomuk.semanticdesktop.org

References

[1] G. Conole, J. Culver, P. Williams, S. Cross, P. Clark and A. Brasher, Cloudworks: social networking for learning design, Proceedings of ascilite Melbourne 2008. Available at: http://www.ascilite.org.au/conferences/melbourne08/procs/conole.pdf (accessed 22 April 2009).
[2] J. Engeström, Why some social network services work and others don't. Or: the case for object-centered sociality (2005). Available at: http://www.zengestrom.com/blog/2005/04/why_some_social.html (accessed 22 April 2009).
[3] K. Knorr Cetina, Sociality with objects. Social relations in postsocial knowledge societies, *Theory, Culture & Society* 14(4) (1997) 1–30.
[4] A. Mathes, Folksonomies – cooperative classification and communication through shared metadata (2004). Available at: http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (accessed 22 April 2009).

[5] T.R. Gruber, Collective knowledge systems: where the Social Web meets the Semantic Web, *Web Semantics: Science, Services and Agents on the World Wide Web* 6(1) (2008) 4–13.

[6] J.G. Breslin and S. Decker, The future of social networks on the internet: the need for semantics, *IEEE Internet Computing* (2007) 86–90.

[7] M. Granovetter, The strength of weak ties, *American Journal of Sociology* 78(6) (1973) 1360–1380.

[8] G. Smith, *Tagging: People-Powered Metadata for the Social Web* (New Riders, Berkeley, CA, 2008).

[9] D. Weinberger, Taxonomies and tags: from trees to piles of leaves, *Esther Dyson's Monthly Report* 23(2) (2005) 1–36.

[10] T. Vander Wal, Folksonomy definition and Wikipedia (2005). Available at: http://www.vanderwal.net/random/entrysel.php?blog=1750 (accessed 22 April 2009).

[11] C. Shirky, Ontology is overrated: categories, links and tags (2005). Available at: http://shirky.com/writings/ontologyoverrated.html (accessed 22 April 2009).

[12] S. Golder and B.A. Huberman, Usage patterns of collaborative tagging systems, *Journal of Information Sciences* 32(2) (2006) 198–208.

[13] C. Marlow, M. Naaman, D. Boyd and M. Davis, Tagging paper, taxonomy, Flickr, academic article, to read, *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (Odense, Denmark, 22–25 August2006). *HYPERTEXT '06* (ACM, New York), 31–40.

[14] H. Halpin, V. Robu and H. Shepherd, The complex dynamics of collaborative tagging, *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada, 08–12 May 2007). *WWW '07* (ACM, New York), 211–220.

[15] P. Mika, Ontologies are us: a unified model of social networks and semantics, *Journal of Web Semantics* 5(1) (2007) 5–15.

[16] C. Cattuto, C. Schmitz, A. Baldassarri, V.D. Servedio, V. Loreto, A. Hotho, M. Grahl and G. Stumme, Network properties of folksonomies, *AI Communications* 20(1) (2007) 245–262.

[17] T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web, *Scientific American* 5(284) (2001) 34–43. Available at: http://www.sciam.com/article.cfm?id=the-semantic-web (accessed 22 April 2009).

[18] T. Berners-Lee, Semantic Web roadmap, World Wide Web Design Issues (2006). Available at: http://www.w3.org/DesignIssues/Semantic.html (accessed 22 April 2009).

[19] H.L. Kim, J.G. Breslin, S.K. Yang, H.G. Kim, int.ere.st: building a tag sharing service with the SCOT ontology, *AAAI 2008 Spring Symposium on Social Information Processing* (Stanford University, CA, USA, 2008).

[20] T.R. Gruber, Ontology of folksonomy: a mash-up of apples and oranges, *International Journal on Semantic Web & Information Systems* 3(2) (2007) 1–11.

[21] T. Berners-Lee, Linked data, World Wide Web Design Issues (2006). Available at: http://www.w3.org/DesignIssues/LinkedData.html (accessed 22 April 2009).

[22] C. Bizer, R. Cyganiak and T. Heath, How to publish Linked Data on the web (2007). Available at: http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial (accessed 22 April 2009).

[23] H.L. Kim, A. Passant, J.G. Breslin, S. Scerri and S. Decker, Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces, *Proceedings of the 2nd IEEE International Conference on Semantic Computing* (Santa Clara, CA, USA, 4–7 August 2008).

[24] H.L. Kim, S. Scerri, J.G. Breslin, S. Decker and H.G. Kim, The state of the art in tag ontologies: a semantic model for tagging and folksonomies, *Proceedings of the International Conference on Dublin Core and Metadata Applications* (Humboldt University, Berlin, 22–26 September 2008).

[25] T. Gruber, 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 22 April 2009).

[26] R. Newman, D. Ayers and S. Russell, Tag ontology. Available at: http://www.holygoat.co.uk/owl/redwood/0.1/tags/ (accessed 22 April 2009).

[27] A. Passant and P. Laublet, Meaning of a tag: a collaborative approach to bridge the gap between tagging and Linked Data, *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008)* (Beijing, China, 2008).

[28] S. Scerri, M. Sintek, L. van Elst and S. Handschuh, NEPOMUK annotation ontology specification (2007). Available at: http://www.semanticdesktop.org/ontologies/nao/ (accessed 22 April 2009).

[29] T. Groza, S. Handschuh, K. Moeller, G.G.L. Sauermann, E. Minack, C. Mesnage, G. MehdiJazayeri and R. Gudjonsdottir, The NEPOMUK Project – on the way to the social semantic desktop, *Proceedings of I-Semantics' 07*, 201–211.

[30] T. Knerr, Tagging ontology – towards a common ontology for folksonomies (2006). Available at: http://code.google.com/p/tagont/ (accessed 22 April 2009).

[31]   F. Echarte, J.J. Astrain, A. Córdoba and J. E. Villadangos, Ontology of folksonomy: a new modelling method, *Proceedings of the Semantic Authoring, Annotation and Knowledge Markup Workshop (SAAKM2007)* (Whistler, British Columbia, Canada, 28–31 October 2007).

[32]   A.Seaborne, and E. Prud'hommeaux, SPARQL query language for RDF (2008). Available at: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/ (accessed 22 April 2009).

[33]   C. Bizer and R. Cyganiak, D2R server: publishing relational databases on the Semantic Web, Poster at the *5th International Semantic Web Conference*, Athens, GA, USA, November 2006.

[34]   H.L. Kim, S.K. Yang, J.G. Breslin and H.G. Kim, Simple algorithms for representing tag frequencies in the SCOT exporter, *IAT 2007: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology* (IEEE Computer Society, Washington, DC, USA, 2007, 536–539).

[35]   U. Bojars, J.G. Breslin, V. Peristeras, G. Tummarello and S. Decker, Interlinking the Social Web with semantics, *IEEE Intelligent Systems* 23(3) (2008) 29–40.

[36]   M. Graves, A. Constabaris and D. Brickley, FOAF: connecting people on the Semantic Web, *Cataloging Classification Quarterly* 43(12) (2007) 191–202.

[37]   P. Miller, R. Styles and T. Heath, Open data commons, a license for open data, *Proceedings of the WWW2008 Workshop on Linked Data on the Web* (Beijing, China, 22 April 2008).

[38]   H. Abelson, B. Adida, M. Linksvayer and N. Yergler, ccREL: The Creative Commons Rights Expression Language (2008). Available at: http://wiki.creativecommons.org/images/d/d6/Ccrel-1.0.pdf (accessed 22 April 2009).

[39]   R. Iannella, Open Digital Rights Language (ODRL) Version 1.1 (2002). Available at: http://www.w3.org/TR/odrl/ (accessed 22 April 2009).

[40]   R. Garca, R. Gil and J. Delgado, A web ontologies framework for digital rights management, *Artificial Intelligence and Law* 15(2) (2007) 137–154.